

INTERNATIONAL  
CONFERENCE

# ADVANCED COMPUTER INFORMATION TECHNOLOGIES



ACIT  
2018



1 - 3 JUNE



Ceske Budejovice,  
CZECH REPUBLIC

## ORGANIZERS :



TERNOPIL NATIONAL  
ECONOMIC UNIVERSITY



Jihočeská univerzita  
v Českých Budějovicích  
University of South Bohemia  
in České Budějovice



## PARTNERS :



MINISTRY  
OF EDUCATION AND SCIENCE  
OF UKRAINE

IEEE computer society

orchitech

TERNOPIL NATIONAL ECONOMIC UNIVERSITY, UKRAINE  
UNIVERSITY OF SOUTH BOHEMIA, CZECH REPUBLIC  
DEGGENDORF INSTITUTE OF TECHNOLOGY, GERMANY

Conference Proceedings

**ADVANCED COMPUTER  
INFORMATION TECHNOLOGIES  
ACIT 2018**

June 1-3, 2018

International Conference

Ceske Budejovice, Czech Republic  
June 1-3, 2018

**ACIT 2018**  
**International Conference**  
**“Advanced Computer Information Technologies”**

**Organized by:**

Ternopil National Economic University, Ukraine  
University of South Bohemia, Czech Republic  
Deggendorf Institute of Technology, Germany

**Partners:**

Ministry of Education and Science of Ukraine  
Czechoslovakia Section of IEEE  
Orchitech

In the Conference proceedings the results of scientific research and development in the following areas: mathematical models of objects and processes, computer network technologies, specialized computer systems, artificial intelligence systems, software engineering, computer and network security, information and analytical support of economic activity, IT law: implementation problems and development prospects, smart grid are published.

The materials are intended for researchers, scientists, teachers, engineers and students.

**Responsible for the issue Mykola Dyvak**

*Approved for publication by the  
Scientific Council of Ternopil National Economic University  
(protocol No 2 on May 16, 2018)*

*Papers are presented in authors edition*

## Conference Chairs



**Andrii Krysovatyi**

Rector of Ternopil National Economic University,  
Professor (Ukraine)



**Tomas Machula**

Rector of University of South Bohemia,  
Professor (Czech Republic)

## Conference Co-Chairs



**Mykola Dyvak**

Dean of Faculty of Computer  
Information Technologies,  
Professor  
(Ternopil National Economic  
University, Ukraine)



**Libor Dostalek**

Head of Institute of Applied  
Informatics,  
Professor  
(University of South Bohemia, Czech  
Republic)



**Wolfgang Dorner**

Head of Institute of Applied  
Informatics,  
Professor  
(Deggendorf Institute of  
Technology, Germany)

## **Dear scientists, participants of the International Conference “Advanced Computer Information Technologies” ACIT’2018!**

The development of civilization causes the rapid development of man-made sphere, which leaves a special trace in the economic, cultural, scientific progress of humanity. The accumulation of knowledge and experience, a stream of new ideas and inventions reveals new opportunities for society and creates new threats. By millennia absorbing huge layers of information, a person wants to separate useful information, use it for analysis and synthesis, find optimal solutions and build effective models.

Among the criteria for the successful existence of companies of the third millennium, there is the one that is based on the collection and processing of data in the monitoring of the environment.

Today, computer technologies with the organization of intelligent computing is experiencing a peak of prosperity. New software products are rapidly emerging, using such advanced technologies and types of tasks, the application of which can achieve significant economic benefits. Computer data processing and intelligent information analysis are now becoming an integral part of the concept of electronic data warehousing and intelligent computing technologies. Dynamic development of the intellectual sphere in any state is a significant factor in increasing the competitiveness of the economy, raising the standard of living of people, improving the quality of technological and environmental safety.

International Conference “Advanced Computer Information Technologies” ACIT’2018 brought together participants from 12 countries: Ukraine, Czech Republic, Germany, Poland, Austria, Russian Federation, Libya, Republic of Korea, India, Lebanon, Iraq, China.

This year, ACIT’2018 is organized by the Faculty of Computer Information Technologies (Ternopil National Economic University, Ukraine), the Institute of Applied Informatics (University of South Bohemia, Czech Republic) and the Institute of Applied Informatics (Deggendorf Institute of Technology, Germany) on the basis of higher education institution which is our partner, at the University of South Bohemia, Ceske Budejovice, Czech Republic. We express our gratitude to our partners for the opportunity to jointly hold this Conference.

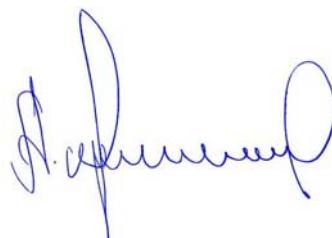
The practice shows that the International Conference “Advanced Computer Information Technologies” enables academics from various universities and IT companies to quickly establish contacts and conclude cooperation agreements.

I am convinced that during the Conference, lively discussions will take place, new ideas and approaches will emerge for solving current problems. I hope that such cooperation will be deepened and improved, gaining new meaning. This meeting of representatives of the scientific elite will become a solid ground for tight cooperation and integration of Ukraine into the world scientific community.

Welcoming the international scientific community, I would like to wish you good health, creative inspiration in conquering new heights of computer science, tireless search, interesting discoveries and implementation of intentions.

Sincerely,

Andriy Krysovatty  
Rector of Ternopil National Economic University



**Dear participants of the International Conference “Advanced Computer Information Technologies” ACIT 2018, it is a great pleasure for us to greet all of you at the Conference**

The first annual All-Ukrainian School-Workshop for Young Scientists and Students “Advanced Computer Information Technologies” was held in May 2011. It was organized by the Faculty of Computer Information Technologies (Ternopil National Economic University, Ukraine), the Association of Computer Information Technologies Specialists and the Council of Young Scientists of Ternopil.

Last year, the VII-th ACIT 2017 has firstly became International Conference. Representatives from three countries took part in the Conference: Ukraine, Poland and Czech Republic. Participants from 23 higher education and scientific institutions of Ukraine were involved.

This year, the VIII-th ACIT 2018 International Conference is organized by the Faculty of Computer Information Technologies (Ternopil National Economic University, Ukraine), Institute of Applied Informatics (University of South Bohemia, Czech Republic) and Institute of Applied Informatics (Deggendorf Institute of Technology, Germany). It is held in the University of South Bohemia, Ceske Budejovice. The University of South Bohemia in Ceske Budejovice has been the center of education, research and science in the southern part of the Czech Republic for more than twenty five years.

We appreciate the great opportunity to jointly hold this International Conference. Eighty six papers were submitted to the Conference. A rigorous review was conducted. We would like to sincerely thank to all reviewers of submitted papers. Their names are listed in the Conference Proceedings. More than 20% of the papers were rejected, sixty eight papers were included in the Conference Program.

Dear participants of the ACIT 2018 International Conference, we wish all of you fruitful work, productive discussions, good inspiration and interesting cooperation in future.

With kind regards,  
Mykola Dyvak, Libor Dostalek, Wolfgang Dorner

## Programme Committee

- Anatoliy Melnyk, Ukraine
- Andreas Berl, Germany
- Andreas Fischer, Germany
- Andriy Pukas, Ukraine
- Frantisek Vacha, Czech Republic
- Iva Dostalkova, Czech Republic
- Krzysztof Gorecki, Poland
- Ladislav Beranek, Czech Republic
- Leonid Lubchik, Ukraine
- Lesia Buyak, Ukraine
- Libor Dostalek, Czech Republic
- Marta Vohnoutova, Czech Republic
- Mikolaj Karpinski, Poland
- Milan Novak, Czech Republic
- Mykola Nikitchenko, Ukraine

- Oleh Berezkyi, Ukraine
- Oleksandr Romaniuk, Ukraine
- Pavel Pudil, Czech Republic
- Peter Hofmann, Germany
- Petro Stakhiv, Poland
- Piotr Szczepaniak, Poland
- Roman Pasichnyk, Ukraine
- Serhii Lupenko, Ukraine
- Vasyl Yatskiv, Ukraine
- Vlasta Svata, Czech Republic
- Volodymyr Stepashko, Ukraine
- Yaroslav Nykolaichuk, Ukraine
- Yevgeniy Bodyanskiy, Ukraine
- Zora Rihova, Czech Republic

### Secretariat

- Ruslan Shevchuk, Ukraine

- Dana Sumerauerova, Czech Republic

### Organizing Committee

- Andrii Kovbasisty, Ukraine
- Andriy Melnyk, Ukraine
- Iryna Darmorost, Ukraine
- Iryna Oliynyk, Ukraine
- Iryna Voytyuk, Ukraine
  
- Andreas Berl, Germany
- Andriy Pukas, Ukraine
- Jan Pour, Czech Republic
- Ladislav Beranek, Czech Republic
- Leonid Lubchik, Ukraine
- Lesia Buyak, Ukraine
- Libor Dostalek, Czech Republic
- Marta Vohnoutova, Czech Republic
- Mikolaj Karpinski, Poland
- Milan Novak, Czech Republic
- Mykola Dyvak, Ukraine
- Mykola Nikitchenko, Ukraine
- Oleh Berezkyi, Ukraine

- Iryna Spivak, Ukraine
- Mykhailo Susla, Ukraine
- Oleksandr Papa, Ukraine
- Svitlana Krepych, Ukraine
- Yurii Maslyiak, Ukraine

### Reviewers

- Oleksandr Romaniuk
- Peter Hofmann, Germany
- Petr Fišer, Czech Republic
- Petro Stakhiv, Poland
- Roman Pasichnyk, Ukraine
- Rudolf Vohnout, Czech Republic
- Ruslan Shevchuk, Ukraine
- Serhii Lupenko, Ukraine
- Vasyl Yatskiv, Ukraine
- Vlasta Svata, Czech Republic
- Volodymyr Stepashko, Ukraine
- Yaroslav Nykolaichuk, Ukraine
- Yevgeniy Bodyanskiy, Ukraine
- Zora Rihova, Czech Republic

## AUTHOR'S INDEX

---

### A

- Alexander Faschingbauer · 280  
 Alexandr Pankratov · 22  
 Alina Davletova · 10  
 Andreas Berl · 280  
 Andrii Bayurskii · 166  
 Andrii Verstiak · 258  
 Andrij Sydor · 110  
 Andriy Dyvak · 54  
 Andriy Goncharenko · 2  
 Andriy Melnyk · 38, 207  
 Andriy Mushak · 6  
 Andriy Pukas · 170, 187  
 Andriy Segin · 10  
 Anna Kovalenko · 22  
 Anna Yaskiv · 128  
 Anton Kalynychenko · 183  
 Antonina Farion · 267  
 Artur Voronych · 123
- 

### B

- Bogdan Markovich · 75  
 Bogdan Maslyiak · 58  
 Bohdan Ihnatiuk · 211  
 Bohdan Tymchyshyn · 207  
 Boris Krulikovskyi · 123
- 

### C

- Christina Sigl · 280  
 Constantine Rudenko · 30
- 

### D

- Daniel Jancarczyk · 102  
 Dmytro Mykhalyk · 42
- 

---

### F

- Ferdinand von Tüllenburg · 285
- 

### G

- Galina Shilo · 14, 183  
 Galyna Grinberg · 34  
 Ganna Plekhanova · 250  
 Georg Panholzer · 285
- 

### H

- Halyna Pidnebesna · 137  
 Haneen A. Barghi · 240  
 Havryshchak · 10
- 

### I

- Igor Boyko · 42  
 Igor Grebennik · 18, 22  
 Igor Yakymenko · 232  
 Inna Urniaieva · 18, 22  
 Iryna Boretska · 83  
 Iryna Darmorost · 26  
 Iryna Mykoliuk · 102  
 Iryna Oliynyk · 46  
 Iryna Pliss · 141, 161  
 Iryna Spivak · 50, 203  
 Iryna Voytyuk · 58, 87, 211  
 Iryna Zhelizniak · 157  
 Iuliia Silantieva · 63  
 Iva Dostalkova · 228  
 Ivan Dychka · 30  
 Ivan Mudryk · 42  
 Ivan Zholubak · 236
- 

### J

- Jakub Geyer · 280  
 Jan Fesl · 97, 224  
 Jan Janeček · 224

Jana Kalova · 106

Jia Lei Du · 285

Jiří Cehák · 224

Jiri Jelinek · 145

Jiří Jelínek · 145

Jiří Pech · 106

Julia Franko · 132

---

### L

Ladislav Beránek · 149

Ladislav Ptáček · 106

Leonid Lyubchyk · 34

Lesia Buiak · 258

Lesia Dubchak · 153

Libor Dostálek · 228

Lidiia Makarova · 199

Lidiya Tereshchenko · 63

Lidiya Tymoshenko · 232

Lyudmyla Honchar · 50

---

### M

Maksym Zaliskyi · 63

Maria Petryk · 42

Marie Doležalová · 224

Mariia Hryhorkiv · 258

Markian Nakonechnyi · 114

Mikolaj Karpinski · 102

Milan Novák · 106

Milos Prokysek · 280

Mohammed Rahma · 236

Mykhailo Kasianchuk · 232

Mykhailo Shpintal · 87

Mykhailo Susla · 38

Mykhailo Tokarchuk · 75

Mykhaylo Petryk · 42

Mykola Dyvak · 46, 50, 54, 58

Mykola Nikitchenko · 175

Mykola Shutko · 63

Mykola Shynkaryk · 26

**N**

- Nadiia Panchuk · 67  
 Nadiia Vasylkiv · 153  
 Natalia Pasichnyk · 38  
 Natalia Porplytsya · 179, 207  
 Natalia Prykhodko · 199  
 Nataliia Furmanova · 14, 183  
 Nataliia Vozna · 110, 123  
 Nataliya Melnyk · 50  
 Nataliya Shakhovska · 157  
 Nina Volianska · 79  
 Nora R. Madi · 240

**O**

- Oksana Bashutska · 67  
 Oksana Huhul · 54  
 Oksana Yaremko · 271  
 Oleg Yurchenko · 128  
 Oleksander Sadovoi · 79  
 Oleksandr Dudnyk · 191, 195  
 Oleksandr Papa · 187  
 Oleksandr Romaniuk · 191,  
   195  
 Oleksandr Viter · 114  
 Oleksiy Sinkevych · 215  
 Olena Boiko · 141  
 Olena Plokha · 250  
 Olena Tsikhanovska · 191  
 Olena Verbova · 153  
 Olena Vynokurova · 161  
 Olexandr Savko · 258  
 Alexandra Viznovych · 75  
 Olha Moroz · 71  
 Orest Ivakhiv · 114  
 Orest Volynskyi · 110

**P**

- Pavlo Kostianoi · 183  
 Petro Humennyi · 110  
 Petro Kostrobij · 75

**R**

- Radim Remeš · 149  
 Radyslav Behota · 114  
 Rashmi L. Jain · 92  
 Rodrigue Elias · 236  
 Roman Krepych · 203  
 Roman Pasichnyk · 38  
 Roman Voliansky · 79  
 Ruslan Avhustyn · 187  
 Ruslan Shevchuk · 232

**S**

- Sachin Jain · 92  
 Seddeq E. Ghrare · 240  
 Sergei Vyatkin · 195  
 Sergey Shekhovtsov · 18  
 Sergiy Prykhodko · 199  
 Serhii Dubovyi · 179  
 Serhii Pavlov · 191  
 Serhii Tryshkaliuk · 87  
 Serhiy Banakh · 271  
 Serhiy Verbovyy · 153  
 Serhiy Yefimenko · 254  
 Stepan Ivasiev · 232  
 Stepan Shkilniak · 175  
 Svitlana Krepych · 166, 203  
 Svitlana Yatsyshyn · 83

**T**

- Tatiana Romanova · 18, 22  
 Tetyana Kulyaba-Kharitonova  
   · 14

**V**

- Vadym Zabchuk · 170  
 Valentyn Tymofieiev · 175  
 Valentyna Panasyuk · 267  
 Valentyna Volkova · 141  
 Valerii Hlukhov · 236

- Vasyl Brych · 50  
 Vasyl Hryhorkiv · 258  
 Vasyl Koval · 118  
 Vasyl Tymchyshyn · 207  
 Vasylyna Melnyk · 26  
 Viktor Kifer · 102  
 Vineet Gokhale · 97, 224  
 Vitalii Dyminskyi · 219  
 Vitalii Smal · 170  
 Vladyslav Holubiev · 211  
 Vlasta Svata · 275  
 Volodymyr Gryga · 123  
 Volodymyr Manzhula · 46  
 Volodymyr Shutko · 63  
 Volodymyr Stepashko · 71,  
   137  
 Volodymyr Tymets · 54  
 Volodymyr Yaskiv · 128

**W**

- Weiwei Zhang · 245

**Y**

- Yaroslav Kaspryshyn · 83  
 Yaroslav Nykolaichuk · 110,  
   123  
 Yaroslav Sokolovskyy · 83,  
   215  
 Yevgeniy Bodyanskiy · 141,  
   161  
 Yevgeniya Sulema · 30  
 Yevhen Kedrin · 87, 187  
 Yuliia Sokhina · 79  
 Yurii Maslyiak · 58  
 Yuriy Batko · 219  
 Yuriy Franko · 132

**Z**

- Zora Říhová · 262

## CONTENTS

### **Mathematical Models of Objects and Processes**

|   |    |
|---|----|
| Andriy Goncharenko. An Entropy Model of the Aircraft Gas Turbine Engine Blades Restoration Method Choice.....   | 2  |
| Andriy Mushak. E-Learning: Application of Compositional and Structural Modeling.....  | 6  |
| Andriy Segin, Alina Davletova, Ihor Havryshchak. Construction of Two-Dimensional Correlation Models in a Cartesian and Spherical Coordinate System.....   | 10 |
| Galina Shilo, Natalia Furmanova, Tetyana Kulyaba-Kharitonova. Software for Tolerance Design of Electronic Devices .....   | 14 |
| Igor Grebennik, Tatiana Romanova, Inna Urniaieva, Sergey Shekhovtsov. Mathematical Model of Balanced Layout Problem Using Combinatorial Configurations.....   | 18 |
| Inna Urniaieva, Igor Grebennik, Tatiana Romanova, Alexandre Pankratov, Anna Kovalenko. Muticriteria Model of Balanced Layout Problem of 3D-Objects .....  | 22 |
| Iryna Darmorost, Petro Stakhiv, Mykola Shynkaryk, Vasylyna Melnyk. Statistical Analysis of Measuring Errors the Pollution of the Atmospheric Bottom Layer by Exhaust Gas .....                          | 26 |
| Ivan Dychka, Yevgeniya Sulema, Constantine Rudenko. A Mathematical Model of Microsurface Normal Distribution for Specular Bidirectional Reflectance Distribution Function.....                          | 30 |
| Leonid Lyubchyk, Galyna Grinberg. Inverse Dynamic Models in Chaotic Systems Identification and Control Problems .....   | 34 |
| Mykhailo Susla, Roman Pasichnyk, Natalia Pasichnyk, Andriy Melnyk. Adjustment of the Model of the Agent-Determinant Type in the Forecasting of Pollution on the Section of the City Road.....           | 38 |
| Mykhaylo Petryk, Dmytro Mykhalyk, Maria Petryk, Igor Boyko, Ivan Mudryk. Modeling of Adsorption and Desorption of Hydrocarbons in Nanoporous Catalyst Zeolite using Nonlinear Langmuir's Isotherm ..... | 42 |
| Mykola Dyvak, Iryna Oliynyk, Volodymyr Manzhula. Design of the Saturated Interval Experiment for the Task of Recurrent Laryngeal Nerve Identification.....  | 46 |
| Mykola Dyvak, Vasyl Brych, Iryna Spivak, Lyudmyla Honchar, Nataliya Melnyk. Discrete Dynamic Model of Retail Trade Market of Computer Equipment in Ukraine.....   | 50 |
| Mykola Dyvak, Volodymyr Tymets, Andriy Dyvak, Oksana Huhul. Methods and Tools for Electrophysiological Monitoring of Recurrent Laryngeal Nerve Monitoring During Surgery on Neck Organs .....           | 54 |
| Mykola Dyvak, Yurii Maslyiak, Iryna Voytyuk, Bogdan Maslyiak. Modified Method of Subtractive Clustering for Modeling of Distribution of Harmful Vehicles Emission Concentrations .....                  | 58 |
| Mykola Shutko, Volodymyr Shutko, Lidiya Tereshchenko, Maksym Zaliskyi, Iuliia Silantieva. Two-Dimensional Spectral Detector for Baggage Inspection X-Ray System.....                                    | 63 |
| Oksana Bashutska, Nadiia Panchuk. Simulation of optimal routes passenger transport.....   | 67 |
| Olha Moroz, Volodymyr Stepashko. Inductive Modeling of Amylolytic Microorganisms Quantity in Copper Polluted Soils .....  | 71 |

|   |    |
|---|----|
| Petro Kostrobij, Bogdan Markovich, Olexandra Viznovych, Mykhailo Tokarchuk. Generalized Transport Equation with Fractality of Space-Time. Zubarev's NSO Method.....   | 75 |
| Roman Voliansky, Oleksander Sadovoi, Yuliia Sokhina, Nina Volianska. Defining of Lyapunov Functions for the Generalized Linear Dynamical Object .....                 | 79 |
| Yaroslav Sokolovskyy, Iryna Boretska, Svitlana Yatsyshyn, Yaroslav Kaspryshyn. Mathematical modeling of deformation-relaxation processes under phase transition ..... | 83 |
| Yevhen Kedrin, Iryna Voytyuk, Serhii Tryshkaliuk, Mykhailo Shpintal. Web Application for Air Quality Monitoring .....   | 87 |

## **Computer Network Technologies**

|   |    |
|---|----|
| Rashmi L. Jain, Sachin Jain. Chain of Clusters for Improving Network Lifetime of Sensor Network ..... | 92 |
| Vineet Gokhale, Jan Fesl. On TCP-Induced Telehaptic Packet Loss and Jitter.....                       | 97 |

## **Specialized Computer Systems**

|   |     |
|---|-----|
| Iryna Mykoliuk, Daniel Jancarczyk, Mikolaj Karpinski, Viktor Kifer. Machine learning methods in ECG classification .....  | 102 |
| Jiří Pech, Milan Novák, Ladislav Ptáček, Jana Kalová. Construction of Vertical Scanner for Laser Analysis of Gel Samples.....   | 106 |
| Nataliia Vozna, Yaroslav Nykolaichuk, Orest Volynskyi, Petro Humennyi, Andrij Sydor. Methods of Crypto Protection of Color Image Pixels in Different Code Systems.....              | 110 |
| Orest Ivakhiv, Markian Nakonechnyi, Oleksandr Viter, Radyslav Behota. Smart Information Gathering Support of Mechatronic System .....   | 114 |
| Vasyl Koval. Algorithms of Landmark Robot Navigation Basing on Monocular Image Processing .....   | 118 |
| Volodymyr Gryga, Yaroslav Nykolaichuk, Nataliia Vozna, Artur Voronych, Boris Krulikovskyi. Development and Research of Conveyor Structures of Binary Number Sorting Algorithms..... | 123 |
| Volodymyr Yaskiv, Anna Yaskiv, Oleg Yurchenko. Synchronous Rectification in High-Frequency MagAmp Power Converters .....  | 128 |
| Yuriy Franko, Julia Franko. Systems for Monitoring Modes and Disturbances in High-Voltage Transmission Lines.....   | 132 |

## **Artificial Intelligence Systems**

|   |     |
|---|-----|
| Halyna Pidnebesna, Volodymyr Stepashko. On Construction of Inductive Modeling Ontology as a Metamodel of the Subject Field.....                                     | 137 |
| Iryna Pliss, Olena Boiko, Valentyna Volkova, Yevgeniy Bodyanskiy. Matrix Deep Neural Network and Its Rapid Learning in Data Science Tasks .....                     | 141 |
| Jiří Jelínek. Classification Model Based on Kohonen Maps .....  | 145 |
| Ladislav Beránek, Radim Remeš. Reasoning With Streamed Information from Unreliable Sources.....   | 149 |
| Lesia Dubchak, Serhiy Verbovyy, Olena Verbova, Nadiia Vasylkiv. Fuzzy Controller of Pathological Conditions Diagnosis Based on Analysis of Cytological Images ..... | 153 |

|   |     |
|---|-----|
| Nataliya Shakhovska, Iryna Zhelizniak. The Associative Rules Constructing on the Example of Patient's Physical Characteristics.....   | 157 |
| Yevgeniy Bodyanskiy, Iryna Pliss, Olena Vynokurova. The Neo-Fuzzy Autoencoder for Adaptive Deep Neural Systems and its Learning ..... | 161 |

## **Software Engineering**

|  |     |
|--|-----|
| Andrii Bayurskii, Svitlana Krepoch. Intelligent System Analyzing Quality of Land Plots .....   | 166 |
| Andriy Pukas, Vitalii Smal, Vadym Zabchuk. Software Based on Blockchain Technology for Consolidation the Medical Data about the Patients Examination .....                             | 170 |
| Mykola Nikitchenko, Stepan Shkilniak, Valentyn Tymofieiev. Satisfiability Problems in Quasiary Program Logics .....  | 175 |
| Natalia Porplytsya, Serhii Dubovyi. Software System for Formation the Composition of Academic Groups (Subgroups) Based on the Diffusion-Like Model .....                               | 179 |
| Nataliia Furmanova, Galina Shilo, Anton Kalynychenko, Pavlo Kostianoi. The Mobile Environment Monitoring System with a Web Interface.....  | 183 |
| Oleksandr Papa, Yevhen Kedrin, Andriy Pukas, Ruslan Avhustyn. Visitors Queue Management Optimization using Web System for Activity Support of the Administrative Services Center ..... | 187 |
| Oleksandr Romaniuk, Oleksandr Dudnyk, Serhii Pavlov, Olena Tsikhanovska. Perspective-Correct Computation Pixels Color for Systems of Three-Dimensional Rendering .....                 | 191 |
| Sergei Vyatkin, Oleksandr Romaniuk, Oleksandr Dudnyk. GPU-Based Rendering for Ray Casting of Multiple Geometric Data.....  | 195 |
| Sergiy Prykhodko, Natalia Prykhodko, Lidiia Makarova. Estimating the Software Size of Open-Source PHP-Based Systems Using Non-Linear Regression Analysis .....                         | 199 |
| Svitlana Krepoch, Iryna Spivak, Roman Krepoch. Research of the Agree of Experts' Evaluations in the Estimation of Software Systems .....   | 203 |
| Vasyl Tymchyshyn, Natalia Porplytsya, Andriy Melnyk, Bohdan Tymchyshyn. Software for Modelling the air Mollution by Vehicles.....  | 207 |
| Vladyslav Holubiev, Bohdan Ihnatiuk, Iryna Voytyuk. Next-Generation Serverless System for Contextual Search Based on Rich Media Content.....   | 211 |
| Yaroslav Sokolovskyy, Oleksiy Sinkevych. Software and Algorithmic Support for representation of CAD models in 2D von Neumann neighborhood .....  | 215 |
| Yuriy Batko, Vitalii Dyminskyi. Fast Contour Tracing Algorithm Based on a Backward Contour Tracing Method .....  | 219 |

## **Computer and Network Security**

|  |     |
|--|-----|
| Jan Fesl, Vineet Gokhale, Marie Doležalová, Jiří Cehák, Jan Janeček. Cloud Infrastructures Protection Technique Based on Virtual Machines Live Migration .....                           | 224 |
| Libor Dostálek, Iva Dostálková . Omnifactor Authentication.....  | 228 |
| Mykhailo Kasianchuk, Igor Yakymenko, Stepan Ivasiev, Ruslan Shevchuk, Lidiya Tymoshenko. The Method of Factorizing Multi-Digit Numbers Based on the Operation of Adding Odd Numbers..... | 232 |

|   |     |
|---|-----|
| Rodrigue Elias, Valerii Hlukhov, Mohammed Rahma, Ivan Zholubak. Hardware Components for Post-Quantum Elliptic Curves Cryptography ..... | 236 |
| Seddeq E. Ghrare, Haneen A. Barghi, Nora R. Madi. New Text Encryption Method Based on Hidden Encrypted Symmetric Key .....              | 240 |
| Weiwei Zhang. A Distributed Security Situation Evaluation Model for Global Network.....   | 245 |

## **Information and Analytical Support of Economic Activity**

|   |     |
|---|-----|
| Ganna Plekhanova, Olena Plokha. Using the Crowdsourcing Online-Platform as IT Tool for Gender Equality Plan Development .....                             | 250 |
| Serhiy Yefimenko. Advances in GMDH-based Predictive Analytics Tools for Business Intelligence Systems .....   | 254 |
| Vasyl Hryhorkiv, Lesia Buiak, Andrii Verstiak, Mariia Hryhorkiv, Olexandr Savko. Management Software Implementing within Ukrainian SMEs: Case Study ..... | 258 |
| Zora Říhová. Integration of Information Systems in Mergers and Acquisition of companies.....  | 262 |

## **IT Law: Implementation Problems and Development Prospects**

|   |     |
|---|-----|
| Antonina Farion, Valentyna Panasyuk. Cybercrimes, Cyber Law and Computer Programs for Security .....  | 267 |
| Oksana Yaremko, Serhiy Banakh. Incitement to Suicide with Social Networks and the Internet: Problems of Criminal Liability in Ukraine ..... | 271 |
| Vlasta Svatá. The Impact of GDPR on IT/IS .....   | 275 |

## **Smart Grid**

|   |     |
|---|-----|
| Christina Sigl, Alexander Faschingbauer, Andreas Berl, Jakub Geyer, Rudolf Vohnout, Miloš Prokýšek. The Role of Smart Meters in P2P Energy Trading in the Low Voltage Grid..... | 280 |
| Ferdinand von Tüllenburg, Jia Lei Du, Georg Panholzer. An Easy-to-use, Scalable and Robust Messaging Solution for Smart Grid Research .....                                     | 285 |

# Mathematical Models of Objects and Processes

# An Entropy Model of the Aircraft Gas Turbine Engine Blades Restoration Method Choice

Andriy Goncharenko

Department of Aircraft Airworthiness Retaining, National Aviation University, UKRAINE, Kyiv, 1 Kosmonavta Komarova ave., email:  
[andygoncharenco@yahoo.com](mailto:andygoncharenco@yahoo.com)

**Abstract:** The paper goal is to investigate the influence of a few specified methods of the gas-turbine engine blades repair upon a choice of a preferable blades geometry restoration technology. There is a scientific proof for the good blades shape restoration method choice that fits the customer needs, being taking into account the subjective preferences of the available technologies and extremizing the preferences uncertainty.

**Keywords:** mathematical model, alternative, economic activity, subjective preference, subjective entropy.

## I. INTRODUCTION

For an active system economic activity management it is used the theory of subjective analysis [1]. A system governed by a decision making person (or a group of such persons in a certain hierarchical structure) is considered as active system since the system is under the managerial influence of the individuals (they are deemed to be reckoned with as the active elements of their own active system). The foundation stone of the subjective analysis theory (the theory of individual subjective preferences optimal distribution upon a certain, taken by an individual into her/his consideration, set of some attainable alternatives) is the entropy paradigm, the so-called Subjective Entropy Maximum Principle (SEMP), which is going to be used in the presented paper as a research tool, although with the application into an objectively existing optimum sphere rather than just to the subjectively preferred matters and only. This creates a background for the information and analytical support of the economic activity.

In aviation industry, in particular, in aeronautical engineering design and its further operation modes, the individuals' subjective preferences play a crucial role. The same is to the aircraft as a whole, as well as to its power plant specifically. Definitely, those individuals' subjective preferences are intruding the fields of both aircraft and their powerplants maintenance and repair as that follows the references [2, 3].

The other area of subjective preferences application here is alternatives in technologies. Concerning an aviation gas-turbine engine (AGTE) repair it deals with the techniques proposed in multiple and developed, described, and discussed in publications of both directions of practitioners/engineers and academicians [4–6].

The objectives of the presented paper are to demonstrate the multi-optimal optimality doctrine newest developments applicability initiated in works [7–11] to the problems of aeroengines technical operation.

The developed herein concept seems promising to the variety of adjacent scientific areas applications, for instance, like for those ones considered and discussed in publications [12–19].

## II. MATHEMATICAL MODELING AND DEVELOPED METHODS

A gas-turbine engine (GTE) blades apparatus is designed on one hand to ensure the required compression of the air supplied with the necessary parameters to the combustion chamber by the compressor part of the engine and on the other hand the blades apparatus of the turbine part allows getting enough work and gas parameters for the compressor driving and aircraft propulsion.

The important thermodynamic parameter here is the polytropic process index  $n$  which magnitude must lie within the very narrow designed value diapason. In operation the blades wear out, their geometrical shape distorts, as a result of this the engine cannot perform the designed work.

The restoration of the blades geometry form has to be made.

### A. Theoretical Problem Setting

Polytropic process index  $n$  approximate mean value, in real engines thermodynamic processes (such as, for instance, expansion of combustion gasses in the cylinder of an internal combustion engine, or in our case, the other examples are compression and expansion of gasses in an AGTE) calculations, can be found from polytropic process equation:

$$pV^n = \text{Const}, \quad (1)$$

where  $p$  is pressure;  $V$  is volume; provided the values of the pressure  $p_1$ ,  $p_2$  and volume  $V_1$ ,  $V_2$  are known at some points 1 and 2 of the process.

Indeed

$$p_1 V_1^n = p_2 V_2^n. \quad (2)$$

$$\ln p_1 + n \ln V_1 = \ln p_2 + n \ln V_2,$$

$$n \ln V_1 - n \ln V_2 = \ln p_2 - \ln p_1,$$

$$n = \frac{\ln p_2 - \ln p_1}{\ln V_1 - \ln V_2}. \quad (3)$$

This simplest method of (1) – (3) for polytropic process index  $n$  determination can be found from practically any reference, guidance or study book on either theoretical or engineering thermodynamics either heat engines.

### B. Optional Functions Entropy Problem Setting

The other concept also proposed in references [7–11] and hereinafter is based upon an optimization principle close to SEMP [1] application in the context close to the described in papers [9, 10].

Let us consider the thermodynamic states 1 and 2 of a gas in polytropic process as some optional states in a certain respect. Thus we come to a multi-optimal problem.

Now, the other sub-problem of the polytropic process given description is to discover the options' objective effectiveness functions related to those two optional states. Let us presuppose the objective effectiveness functions for the considered two-optimal problem of the polytropic process considered description are  $\ln V_1$  and  $\ln V_2$ . This might be reasonably natural with regards to apparent perception of the obvious quantitative characteristic of the existing reality.

With the use of the supposed multi-optimal optimality, likewise in subjective analysis [1, 10] conditional optimality of the individual's subjective preferences distribution, with extremizing subjective entropy, that is applying the doctrine analogous to SEMP concept, we have the right to write down the postulated functional in the view of:

$$\Phi_h = - \sum_{i=1}^2 h_i(V_i) \ln h_i(V_i) + n \sum_{i=1}^2 h_i(V_i) \ln V_i + \gamma \left[ \sum_{i=1}^2 h_i(V_i) - 1 \right], \quad (4)$$

where  $h_i(V_i)$  are specific hybrid-optimal effectiveness functions, similar to the preferences functions of [1, 10], however in this problem setting the assumed specific hybrid-optimal effectiveness functions  $h_i(V_i)$  are not relating with anybody's preferences or choice;  $\gamma$  is normalizing coefficient (function).

The first member of expression (4) is the hybrid-optimal effectiveness functions entropy (like subjective entropy of the preferences).

The necessary conditions for the functional (4) extremum existence:

$$\frac{\partial \Phi_h}{\partial h_i(V_i)} = 0, \quad (5)$$

yield

$$\frac{\partial \Phi_h}{\partial h_i(V_i)} = -\ln h_i(V_i) - 1 + n \ln V_i + \gamma = 0, \quad \forall i = \overline{1, 2} \quad (6)$$

This inevitably means in turn

$$\ln h_1(V_1) - n \ln V_1 = \gamma - 1 = \ln h_2(V_2) - n \ln V_2. \quad (7)$$

From where

$$\ln h_1(V_1) - \ln h_2(V_2) = n(\ln V_1 - \ln V_2). \quad (8)$$

And

$$n = \frac{\ln h_1(V_1) - \ln h_2(V_2)}{\ln V_1 - \ln V_2}. \quad (9)$$

### III. ANALYSIS

#### A. Polytropic Equation Setting

Thus, we have got a parallel result to the law of subjective conservatism if the values of parameters  $n$ ,  $V_1$ , and  $V_2$  are given.

In case as in work [9]:

$$h_1(V_1) = xp_2, \quad h_2(V_2) = xp_1, \quad (10)$$

where  $x$  is unknown, uncertain multiplier in type of the Lagrange one, we obtain with the help of the procedure considered through (4) – (10) the needed polytropic process index (3).

Indeed, substituting equations (10) for their values into expression (9) it yields

$$n = \frac{\ln \frac{xp_2}{xp_1}}{\ln V_1 - \ln V_2}, \quad (11)$$

finally, formula (3).

The sense of the uncertain multiplier  $x$  becomes obvious with the use of the normalizing condition of the initial functional (4). That is

$$xp_1 + xp_2 = 1. \quad (12)$$

Hence,

$$x = \frac{1}{p_1 + p_2}. \quad (13)$$

Remarkable here is that the multi-optimal hybrid function has the view of

$$n \frac{\sum_{i=1}^2 p_i \ln V_i}{\sum_{i=1}^2 p_i} \quad \text{or} \quad n \frac{\sum_{i=1}^2 p_{\bar{i}} \ln V_i}{\sum_{i=1}^2 p_i}, \quad (14)$$

where subscript  $\bar{i}$  means: “pertaining not to the  $i^{th}$  but to the other option of the two-optimal situation”.

Moreover, in case expressed with (4) – (14) the sought polytropic process index  $n$  has been got for the given values of hybrid-optimal effectiveness functions  $h_i(V_i)$  and at this the found value of  $n$  can make the hybrid-optimal effectiveness functions of  $h_i(V_i)$  also be optimal for the objective functional (4).

#### B. Aeroengine blades alternative restoration technique problem solution on the subjective entropy paradigm basis

Now, in order to retain polytropic process index within the required designed value interval, so that to ensure the desired engine performance, periodical restoration of the specified GTE blades apparatus geometry has to be executed.

Supposedly, there are competing methods. These are, for example, plasma (Pl), laser (La), and electro-arc (El). Each of which are the alternatives for a GTE repair plant to implement. Here we have a three-alternative problem. Then, there are corresponding subjective effectiveness functions: Pl( $\cdot$ ), La( $\cdot$ ), and El( $\cdot$ ) that have relations to each of the being considered alternatives.

Let us apply although simplified, rough, however possible model for the objective effectiveness functions  $Pl(\cdot)$ ,  $La(\cdot)$ , and  $El(\cdot)$ .

The results of the numerical modeling (in conditional units) are shown in Fig. 1–3.

In the considered example the objective effectiveness functions  $Pl(\cdot)$ ,  $La(\cdot)$ , and  $El(\cdot)$  have five independent variables. But in actual result presented in Fig. 1 only one: # 4 (productivity of the alternative technology:  $P$ ) is being variated. The rest of the parameters are the corresponding constant values accepted for the: 1) thickness of the metal layer welded onto the blades prepared surface; 2) thickness of the metal layer removed out from the blades bodies down to the nominal size; 3) number of blades undergoing the treatment; and 5) is the number of the laborers supposed to be involved into the alternative technological process.

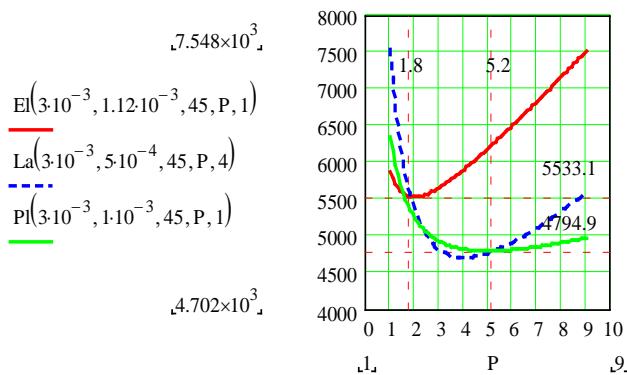


Fig.1. Objective effectiveness functions related to corresponding alternatives

### C. Results of the Numerical Experiment

For the available three alternatives: plasma, laser, and electro-arc, all the described approach variables may also be functions of their arguments. Even with the simplified set of the independent variables the situation depending upon the arguments combination remains uncertain. The inter-influence of the parameters can lead to variants when at some circumstances it is hard to give the preference to a specific alternative.

The preferences obtained by the objective functional treatment like (4) – (6) are visible in Fig. 2.

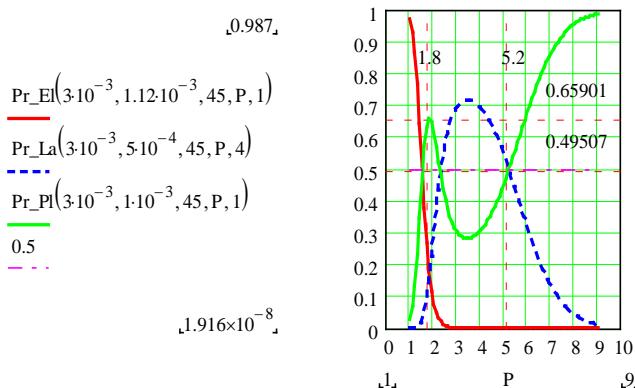


Fig.2. Subjective preferences of corresponding alternatives

In Fig. 2 it is depicted, with the  $Pr$  characters, the subjective preferences functions distributed optimally, accordingly with the procedures of (4) – (6) [1], upon the set of the alternative AGTE blades restoration technologies:  $Pl(\cdot)$ ,  $La(\cdot)$ , and  $El(\cdot)$ , correspondingly.

For the preferences, their four arguments out of the five introduced are fixed exactly as for the corresponding effectiveness functions (see designations in Fig.1 and 2).

Subjective entropy of preferences is presented in Fig. 3.

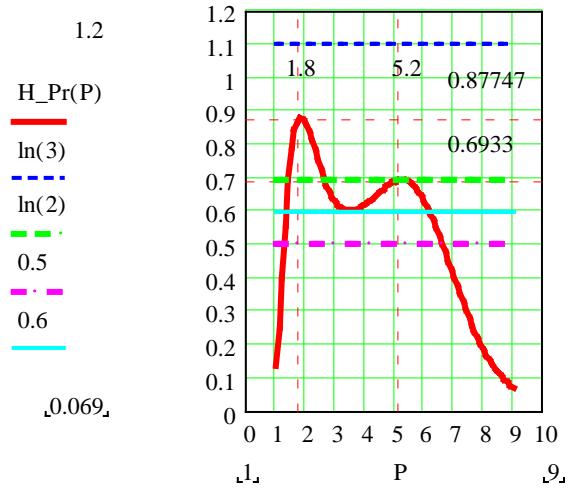


Fig.3. Subjective entropy of corresponding alternatives preferences

From Fig. 3 it is visible that there are areas with respect to the productivity  $P$  of the blades restoration methods from approximately 1.5 up to 6.5 kg/h where it is almost impossible to make a decision concerning which technology is better to apply (see and compare multi-alternativeness knots of the individual subjective preferences, noticeable in Fig. 2 at  $P = 1.8$  and  $5.2$  kg/h, with the subjective entropy climaxes appeared in Fig. 3). It might be suspected because of the effectiveness functions positioning (also see Fig. 1 at those values).

The additional information is required to decrease the subjective entropy. In the framework of the presented model and developed doctrine the mathematical expressions constructed and the major parameters being considered will undoubtedly lead to factual realizations. This process of the modifications creation is a challenging task; and it is absolutely clear that the necessary additional information needed to decrease the uncertainty of the alternative AGTE blades restoration technologies (subjective entropy of the available preferences) lies in the sphere of the technological processes intrinsic values selected each time by the researcher to compare the alternatives. Therefore, the essential parameters of the processes models, as well as the models' own plausible mathematical constructions, for every stated problem setting, embody the significant information that finally decreases the entropy value.

Thus, if we suppose existence of the different thresholds of the entropy: 0.5; 0.6; 0.7 (see Fig. 3) for the corresponding choice, then the subjective entropy as the measure of uncertainty will be higher or lower than mentioned thresholds. More to the point, there are two extrema at

$P = 1.8$  and  $5.2 \text{ kg/h}$ ; the latter says of practically two-alternative situation since the uncertainty almost coincides with the two-alternative situation uncertainty maximal value  $\ln 2$  (see Fig. 3).

#### IV. CONCLUSION

In view of the stated problem, the developed doctrine allows formulating the following new scientific results.

The proposed approach gives a possibility of the alternative aviation gas-turbine engine blades restoration technologies comparison based upon the application of the entropy extremization principle. The doctrine implementation reflects the results of the goal achieving for both objectively existing and subjectively preferable optimums, as well as for their combinations.

Thus, for the absolutely objectively existing polytropic process, the pressure in polytropic process is the optimal hybrid-optimal function, measured with certain units, of the “*logarithmic measureless volume*”. Furthermore the polytropic process index is obtained on the basis of the multi-optimal entropy conditional optimality doctrine rather than on the absolutely thermodynamic derivations.

For the subjective component the preferences functions allow the alternatives assessing with the uncertainty measure.

In further research it should be considered some other effectiveness functions and their variables, as well as found more theoretical results and applicable areas of the hybrid-optimality doctrine.

#### REFERENCES

- [1] V. Kasianov, Subjective Entropy of Preferences. *Subjective Analysis: Monograph*, Institute of Aviation Scientific Publications, Warsaw, Poland, 2013, 644 p. ISBN: 978-83-63539-08-5.
- [2] T. W. Wild and M. J. Kroes, Aircraft Powerplants, McGraw-Hill, Education, New York, NY, USA, 2014, 756 p.
- [3] M. J. Kroes, W. A. Watkins, F. Delp, and R. Sterkenburg, Aircraft Maintenance and Repair, McGraw-Hill, Education, New York, NY, USA, 2013, 736 p.
- [4] Y. A. Tamarin, Protective Coatings for Turbine Blades, ASM International, Materials Park, Ohio, USA, 2002, 247 p. ISBN: 0-87170-8809-759-4.
- [5] K. J. Pallos, Gas Turbine Repair Technology, GE Energy Services Technology, GE Power Systems, Atlanta, USA, 2001, 30 p. (GER-3957B).
- [6] S. Dmitriev, A. Koudrin, A. Labunets, and M. Kindrachuk, “Functional coatings application for strengthening and restoration of aviation products,” *Aviation*, vol. 9, no. 4, pp. 39-45, 2005.
- [7] A. V. Goncharenko, “Optimal UAV maintenance periodicity obtained on the multi-optimal basis,” in *Proceedings of the IEEE 4th International Conference on Actual Problems of Unmanned Aerial Vehicles Developments (APUAVD)*, IEEE, Kyiv, Ukraine, October 2017, pp. 65-68.
- [8] A. V. Goncharenko, “Optimal controlling path determination with the help of hybrid optional functions distributions,” *Radio Electronics, Computer Science, Control*, vol. 1(44), pp. 149-158, 2018.
- [9] A. V. Goncharenko, “Aeronautical and aerospace materials and structures damages to failures: theoretical concepts,” *International Journal of Aerospace Engineering*, Article ID 4126085, 7 pages, 2018 <https://doi.org/10.1155/2018/4126085>.
- [10] A. V. Goncharenko, “Aircraft operation depending upon the uncertainty of maintenance alternatives,” *Aviation*, vol. 21, no. 4, pp. 126-131, 2017. <https://doi.org/10.3846/16487788.2017.1415227>
- [11] A. V. Goncharenko, “Multi-Optional Hybrid Effectiveness Functions Optimality Doctrine for Maintenance Purposes,” in *Proceedings of the 14th IEEE International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET-2018)*, IEEE, Lviv-Slavskie, Ukraine, February 2018, pp. 771-775.
- [12] M. Dyvak, V. Tymets, and V. Brych, “Improving the Effectiveness of Electrophysiological Monitoring of the Recurrence Laryngeal Nerve During Surgery on Neck Organs,” in *Proceedings of the IEEE 14th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET-2018)*, Lviv-Slavskie, Ukraine, February 2018, Paper No. 153, Paper ID 387.
- [13] M. Dyvak, I. Oliynyk, Y. Maslyiak, and A. Pukas, “Static interval model of air pollution by motor vehicles and its identification method,” in *Proceedings of the IEEE 14th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET-2018)*, Lviv-Slavskie, Ukraine, February 2018, Paper No. 176, Paper ID 364.
- [14] O. Zaporozhets, V. Tokarev, and K. Attenborough, Aircraft Noise. Assessment, Prediction and Control, Tailor and Francis, Glyph International, 2011, 480 p.
- [15] O. Solomentsev, M. Zaliskyi, and O. Zuiev, “Estimation of quality parameters in the radio flight support operational system,” *Aviation*, vol. 20, no. 3, pp. 123-128, 2016.
- [16] T. Shmelova, Y. Sikirda, N. Rizun, A.-B. M. Salem, and Y. N. Kovalyov, Socio-Technical Decision Support in Air Navigation Systems: Emerging Research and Opportunities, *International Publisher of Progressive Information Science and Technology Research*, Pennsylvania, USA, 2017, 264 p.
- [17] V. Infante, J. M. Silva, M. de Freitas, and L. Reis, “Failures analysis of compressor blades of aeroengines due to service,” *Engineering Failure Analysis*, vol. 16, no. 4, pp. 1118-1125, 2009.
- [18] A. Palacios, A. Martínez, L. Sánchez, and I. Couso, “Sequential pattern mining applied to aeroengine condition monitoring with uncertain health data,” *Engineering Applications of Artificial Intelligence*, vol. 44, pp. 10-24, 2015.
- [19] A. Innocenti, L. Marini, and C. Carcasci, “Effects of Upgraded Cooling System and New Blade Materials on a Real Gas Turbine Performance,” *Energy Procedia*, vol. 101, pp. 1135-142, 2016.

# E-Learning: Application of Compositional and Structural Modeling

Andriy Mushak

Department of Economic Cybernetics and Informatics, Ternopil National Economic University, UKRAINE, Ternopil, 3 Peremohy sq.,  
email: [andriy\\_mushak@hotmail.com](mailto:andriy_mushak@hotmail.com)

**Abstract:** It is proposed to use the elements of compositional and structural modeling (CSM) to solve the problems related to the development of web-based interactive multimedia software applications for distance learning and training. When developing learning and training software with CSM approach, the mathematical model of learning and training system with problem solution finding algorithm developed makes it possible to implement a methodology of using interactive multimedia means in distant learning and training. The approach proposed to develop distance learning and training courses increases the technological flexibility of learning and training processes. Based on the approach proposed, a number of learning and training courses has been already developed.

**Keywords:** e-learning, distance learning, compositional and structural modeling, multimedia programs for educational purposes, the application software system.

## I. INTRODUCTION

The level of education of the information society is an essential element of its development. Education is referred to strategically important areas of implementation of telecommunication and information technologies in Ukraine, because the level of knowledge of each person creates the foundation for the development of our state. The continuous development of information systems and technologies, which is accompanied by their implementation in everyday life, the gradual transition to high-tech production and, in general, an increase in the rhythm of life, require from each individual continuous improvement of the acquired level of knowledge and mastering of substantially new knowledge. In other words, it is about the need for lifelong learning.

To solve this global task, all available means are acceptable. These include both classical and modern forms of learning [1-4]. Here the important place belongs to distance learning through global computer networks with using communication and information technologies, because it has a number of very important advantages in the context of today's situation compared then traditional:

- satisfies the individual choice of training trajectory: mode, time and speed;
- satisfies unimpeded access of students to training materials;
- satisfies constant contact with the tutor: the student can contact the tutor at any time and ask for help;

- provides the opportunity to hire foreign tuto. Using the Internet, they have the opportunity to simultaneously train all those who want from different parts of the world. Tutors do not need to move from country to country to hold classes;
- satisfies constant communication between students for the purpose of discussion current issues during the processing of the training material, and for contacting for mutual interests;
- brings economic benefits. An increase in the number of students does not require significant additional costs. In this sentence and further, the word "student" should be understood by everyone who is studying on the distance learning program (in particular, students from secondary schools, university students, retraining people, and, in general, raise their level of knowledge, etc.).

Today distance learning technologies are developing intensively. Native and foreign scientists have significant achievements in the study of the methodology of creating interactive distance multimedia educational programs: O.M. Dovgyallo, V.N. Kukharenko, M.I. Zhaldak, V.V. Lapinsky, V.M. Tomashevsky, P. Commerce and others [2-7]. They have a number of significant results in developing models, methods and technologies for distance learning, but a number of problematic issues still need to be addressed.

The primary problem that is currently being addressed is the construction of a methodology, which would include, on the one hand, methods of increasing the productivity during the simulation of distance learning processes, and, on the other, methods for research and analysis of the effectiveness of the type of study.

In the works [8, 9] proposed models and methods that, in general, allow us to talk about the creation of elements of the methodology of constructing distance learning courses (DLC). It is in this, as well as the possibility of using the proposed approach to various technologies for designing and creating software (including object-oriented) and is the relevance of the study.

## II. PURPOSE AND TASKS OF THE INVESTIGATION

The purpose of the research is to develop elements of the methodology for constructing interactive remote multimedia educational programs.

In accordance with the stated goal in the article put and solve such problems:

- to examine the principles of compositional and structural modeling technology;

- to construct a mathematical model of the teaching system;
- to study the application of compositional structural modeling technology elements (CSM-technology) in DLC;
- to realize the method of service by means of interactive multimedia in distance learning.

### III. ELEMENTS OF COMPOSITIONAL AND STRUCTURAL MODELING TECHNOLOGY

Consider the elements of CSM-technology in order to use some of them to build a DLC [10, 11]. This technology allows to increase the productivity of the developers of application software systems (ASS) (which includes also education software systems), improve the quality and reliability of such systems through the development of unified mechanisms, language models, methodologies for building ASSs.

The monolithic way of designing applications, and, consequently, their next programming, is characteristic of 1st generation software. The complexity and rising cost of ASS allows us to conclude that this method is inadequate.

Technological programming principle is the modularity in which the program is designed as a chain of components ("bricks"), called modules. Each of these "bricks" acts as a separate program unit. It is designed autonomously, autonomously programmed and tested, used in a wide variety of programs as an integral part, when only for its functional purpose, the module meets the needs. Modularity provides a structural adaptation of the algorithm to the problem to be solved, it can be connected to new "bricks", modify and renew the old ones up to the design of a completely new algorithm.

The conceptual foundations of macromodule programming are considered. In particular, the description of syntactic models of the languages of the macromodual programming environment and the description of the verification of composite schemes should be used to achieve the set goals.

When constructing DLC for CSM-technology there is a series of optimization problems that are generally characteristic of different stages of ASS development. In particular, the problem of choosing an optimal algorithm in a given set of competing algorithms under various practically important assumptions about the properties of the latter can be formulated in this way.

Let the set of algorithms  $A_j \subseteq A$  is defined for solving the problem  $z_j \in Z$ , by means of which this problem can be solved. Algorithms  $A_{ji} \in A$  are in accordance with the sequence of characterizing their parameters  $\alpha_{ji} = \{\alpha_{ji}^k : k = 1, \dots, q\}$ . In the set  $A_j$  it is necessary to choose the algorithm  $A_{je}$  such that

$$\phi(\alpha_{je}, \gamma_{je}) = \text{ext}_i \phi(\alpha_{ji}, \gamma_{ji}) \quad (1)$$

with some limitations on  $\alpha_{ji}$ .

This problem in the general case is a complex task of

multicriteria optimization and, taking into account the needs of the practice of constructing a DLC, may acquire (by equating individual parameters  $a_k$  to zero) different partial formulas of the type:

– among all the algorithms  $A$  to find at least one algorithm that can solve this problem;

– in the set of algorithms  $A$  to find the most effective one by one indicator, for example, for the speed to solve this problem, and so on.

Practical interest is the search for such an algorithm, all indicators which would most closely match the requirements of the user in solving this problem.

### IV. APPLICATIONS OF CSM-TECHNOLOGY ELEMENTS FOR DISTANCE LEARNING

Demonstration of the use of elements of CSM-technology in the DLC may be, in particular, a number of examples of serving the technology under consideration for the implementation of learning tasks.

A mathematical model of the learning system is constructed by introducing so-called expansion functions  $G_{\langle x^*, y^* \rangle}$  in the state space, defining the operations of their sum  $G_{\langle x^*, y^* \rangle} + H_{\langle u^*, v^* \rangle}$ , the product  $G_{\langle x^*, y^* \rangle} \circ H_{\langle u^*, v^* \rangle}$  and the complex function  $R_{\langle r,s \rangle}(x)$ , which is given by a recursive scheme:

$$R_{\langle r,s \rangle}(x) = G_{\langle x^*, y^* \rangle}(x) | R_{\langle r,s \rangle}(x) + H_{\langle u^*, v^* \rangle}(x) | R_{\langle r,s \rangle}(x) \quad (2)$$

$$\circ H_{\langle u^*, v^* \rangle}(x) | H_{\langle u^*, v^* \rangle}(x) \circ R_{\langle r,s \rangle}(x)$$

The problem on a mathematical model  $C$  is called a pair of states  $\langle x_0, y_0 \rangle$ .

The complex function  $R_{\langle r,s \rangle}(x)$  is called the solution of the problem  $\langle x_0, y_0 \rangle$  on the model  $C$ , if the following conditions are satisfied:

- 1)  $x_0$  belongs to the area of function definition  $R_{\langle r,s \rangle}(x)$ , i.e.  $x_0 \in Z = \{z : z \geq r\}$ .
- 2)  $R_{\langle r,s \rangle}(x_0) \geq y_0$ .

It is shown that in order for the function  $R_{\langle r,s \rangle}(x)$  to be a solution of the problem  $\langle x_0, y_0 \rangle$ , it is necessary and sufficient that the condition  $x_0 \geq y_0 s \vee r$ .

The algorithm for constructing a solution of the problem  $\langle x_0, y_0 \rangle$  on the  $C$  model reduces to the execution of such a sequence of steps:

- 1) put

$$W_1 = \left\{ G_{\langle x(j), y(j) \rangle}^1(x) : G_{\langle x(j), y(j) \rangle}^1(x) \in \mathfrak{I}, x(j) \leq x_0, j = 1, \dots, m_1 \right\}$$

$$2) \text{ let } G_{\langle x,y \rangle}^1(x) = \left( \sum_{j=1}^{m_1} G_{\langle x(j), y(j) \rangle}^1 \right)(x)$$

and  $d_1 = G_{\langle x,y \rangle}^1(x_0)$ , where

$$x = \bigvee_{j=1}^{m_1} x(j), \quad y = \bigvee_{j=1}^{m_1} y(j);$$

3) organize the iterative process of constructing  $W_i$  sets in this way:

$$W_i = \left\{ G_{\langle x(j), y(j) \rangle}^i(x) : G_{\langle x(j), y(j) \rangle}^i(x) \in \mathfrak{I} \setminus \bigcup_{l=1}^{i-1} W_l, x(j) \leq d_{i-1}, j = 1, \dots, m_i \right\}$$

$$G_{\langle x,y \rangle}^i(x) = \left( \sum_{j=1}^{m_i} G_{\langle x(j), y(j) \rangle}^i \right)(x) \text{ and}$$

$$d_i = G_{\langle x,y \rangle}^i(d_{i-1}), \quad x = \bigvee_{j=1}^{m_i} x(j), \quad y = \bigvee_{j=1}^{m_i} y(j);$$

4) the iterative process stops on condition  $W_i = \emptyset$ ;

5) if  $W_i = \emptyset$  on the  $k+1$  step of the iterative process, then put

$$R_{\langle r,s \rangle}(x) = \left( \prod_{i=1}^k G_{\langle x,y \rangle}^i \right)(x) \quad (3)$$

The statement and the given algorithm allow to strictly solve tasks of control of the correctness of tasks execution, in particular, the construction of schedules of functions. This is achieved by determining the conditions under which the given graph can be constructed, as well as all possible ways of constructing the graph, which we obtain as different (relative to the commutativity) of the solution of the problem on the formal model.

We will describe the details of the program implementation of one example, this is the construction of graphs of functions. The essence of this task is to enable the student to master the construction of graphs of such functions, which are a composition of other (simpler functions). It is obvious that modules for this task will be represented in the context of CSM-technology. These are programs that realized the construction of a graph of a simple function or their composition. Modules are also a program that allows graphs to be displayed as sets of pixels, and also coordinate plane with the necessary infrastructure.

The requested software is written in Java. Each module of the software system is an applet. For simplicity, an integrated programming environment Borland JBuilder was used.

Particular attention is paid to organizing the code. Note that modules are not static units. Therefore, the nuances of data transfer between them are noted during the application of the AppletContent interface. It is known that the interaction between applets located on the same HTML page only involves calling from one of the applets (applet client) to the method specified in another applet (applet server). In our case, the applet server is an applet "Coordinate system". The rest of the applets are client

applets. Each client applet has an approach to the method AddPoint(int Value\_of\_Function, int red\_Ingrad, int green\_Ingrad, int blue\_Ingrad) of the applet server. This is realized using the applet context.

```
appletServer=getAppletContext().getApplet("CoordinatePlane");
((Applet1)appletServer).AddPoint(y,red_Ingradient,
green_Ingradient,blue_Ingradient);
```

The AddPoint method adds a new point to the value array. The above fragment of the listing states that this method is parametrized, not only the value of the function is transmitted, but also the values of the three values of the type int, constituents of the color, which will display the graph. After the formation of an array of values, the paint() method, which is blocked by us, is executed. This brings up a new graph. Construct graphs of the functions  $y = -(f(x))$ ,  $y = f(|x|)$ ,  $y = |f(x)|$ ,  $y = |f(|x|)|$  and  $y = f_1(x) + f_2(x)$  involves getting one (Value\_of\_Function\_Array\_Second[]) or two arrays in the corresponding applet (Value\_of\_Function\_Array\_First[], Value\_of\_Function\_Array\_Second[]) values of the function for its next (their) processing. To do this, the following methods are defined in the "Coordinate system" applet.

```
public int[]
get_Value_of_Function_Array_Second() {
    return Value_of_Function_Array_Second;
}
```

```
public int[]
get_Value_of_Function_Array_First() {
    return Value_of_Function_Array_First;
}
```

The nature of these methods is trivial. They return the value function arrays. Next, using the context of the applet, this data is reading, for example

```
appletServer=getAppletContext().getApplet("CoordinatePlane");
Value_of_Function_Array_Local=((Applet1)
appletServer).get_Value_of_Function_Array_Second();
```

after which the processing of the array

Value\_of\_Function\_Array\_Local

is in progress.

Practical implementation ends with consideration of the use of interactivity in the DLC "Placing Productive Forces of

Ukraine", which was used to train students at the International University of Finance. The "Course of Communication and Information Technologies", "Interactive Training Program for Teachers Using Telematics in Distance Learning" and others were built.

It is noted, in particular, how and by what means of interactive multimedia the lecture material of the course is presented, how the control of the knowledge received by the student is organized and what are the features of the course design.

## V. CONCLUSION

Applying the method of using interactive multimedia tools in the DLC, an approach has been developed on the use of CSM-technology in the development of DLC fragments. The mathematical model of the teaching system with the algorithm of finding solutions of problems in the construction of teaching programs for CSM-technology is proposed.

The approach to constructing interactive teaching programs with the help of CSM-technology is developed. It is demonstrated on a concrete example of the application of CSM-technology in distance learning courses. A number of problems that arise during the learning process are solved, in particular during the construction of graphs of elementary functions.

Approaches to modeling different processes of distance learning can be the basis of the tool environment for supporting the development of distance learning courses, taking into account the whole complex of methodological problems that arise in this case.

## REFERENCES

- [1] Belov, V.N., Dovgyallo, A.M. "Principles of organization and results of experimental testing of a package of subprograms oriented to the production of interactive learning programs". *Control systems and machines*, no. 1, 1978, pp. 41-47.
- [2] Dovgyallo, A.M., Yushchenko, Ye.L. "Learning systems of the new generation". *Control systems and machines*, no. 1, 1988, pp. 83-86.
- [3] Kukharenko, V.M. *Distance Learning: Terms of Use*. Ukraine, Kharkiv: Kharkiv Polytechnic Institute, National Technical University, 2001, P. 212.
- [4] Kukharenko, V.M. *Distance Learning: Terms of Use*. Ukraine, Kharkiv: Torsinh, 2002, P. 298.
- [5] Tomashevskij, V.M., Zgurovskij, M.Z. *System modelling*. Ukraine, Kyiv: BHV, 2005, P. 352.
- [6] Kommers, P., Semerling, M. *Information and Communication Technologies for Secondary Education*. Russia, Moscow: Publishing House «Obuchenie-Servis», 2005, P. 264.
- [7] Mushak, A., Provtar, O. "Distance learning: from model building to software code generation". *Digest of the Ternopil National Technical University*, no. 1, 2013, pp. 107-115.
- [8] Mushak, A.Ya. Computer modeling of distance learning processes in Internet technologies (PhD Thesis), Kyiv: Glushkov Institute of Cybernetic of NAS of Ukraine, 2004, P. 40.
- [9] Mushak, A. "Methodology for using interactive multimedia in distance learning". Proceedings of the *Problems of introduction of information technologies in economy and business: International scientific and practical conference (Ukraine, Irpin, April 18-22, 2000)*, Irpin: Problems of introduction of information technologies in economy and business, 2000, pp. 319-320.
- [10] Mushak, A. "Use of communication and information technologies by teachers". Proceedings of the *Distance Education: Open and Virtual Environments: VII International Conference on Distance Education (Russia, Moscow, June 12-16, 1999)*, Moscow: Distance Education: Open and Virtual Environments, 1999, pp. 190-197.
- [11] Mushak, A. "Using the interactive multimedia in distance course «Communication and information technologies (CIT-course)»". Proceedings of the *Seminar about Computers in School POSKOLE'99* (Czech Republic, Lazne Sedmihorky, November 23-26, 1999), Czech Republic, Lazne Sedmihorky: Seminar about Computers in School POSKOLE'99, 1999, pp.119-121.

# Construction of Two-Dimensional Correlation Models in a Cartesian and Spherical Coordinate System

Andriy Segin<sup>1</sup>, Alina Davletova<sup>1</sup>, Ihor Havryshchak<sup>2</sup>

1. Department of Specialized Computer Systems, Ternopil National Economic University, UKRAINE, Ternopil, 8 Chehova str., email: [ase@tneu.edu.ua](mailto:ase@tneu.edu.ua), [a90f@meta.ua](mailto:a90f@meta.ua)

2. Department of the Ukrainian Language, I. Horbachevsky Ternopil State Medical University, UKRAINE, Ternopil, 1 Voli sq.  
e-mail: [havruchukii@tdmu.edu.ua](mailto:havruchukii@tdmu.edu.ua)

**Abstract:** This paper presents a methodology for constructing and visualizing correlation models of two-dimensional signals in a rectangular and spherical coordinate system.

**Keywords:** correlation models, spherical coordinate system, two-dimensional models.

## I. INTRODUCTION

The mathematical apparatus of correlation analysis in the Cartesian coordinate system remains a powerful tool for researching technological processes and natural phenomena [1]. Areas of its application are extremely diverse both in the fundamental directions of science and applied science. Correlation analysis is successfully used in the study of processes in atomic and nuclear physics, energy, electronics, radio engineering, astronomy, astrophysics, economics and other fields of science [2-4].

Correlation models give an opportunity to investigate both one-dimensional and multidimensional deterministic and stochastic processes [5]. It is clear that the correlation analysis is not universal, has its own limitations and possibilities of use for a certain class of tasks. However, this is a convenient and reliable method for solving a wide range of tasks, which needs further development and improvement for its greater efficiency and extending its use.

At the present stage of microprocessor development and computer technology, new perspectives of correlation models development are opened, which will allow them to be used in real-time systems and create more complex models.

The article presents mathematical expressions for constructing correlation models in a spherical coordinate system and proposes to take into account the effect of "aging" information and transport delays in correlation models. The construction of correlation models in a spherical coordinate system opens up new prospects for their use in the study of processes and phenomena, which are simpler described in spherical coordinates than in Cartesian space. Such processes include the determination of orbits and figures of celestial bodies, the determination of objects' movement in space, radar, and propagation of waves of various nature in space, trajectories of motion of elemental particles in atoms, rotating processes in mechanics, determination of precise coordinates in cartography, and many others.

The article presents the first theoretical stage of the study of the satellite's interconnection in a stationary orbit with points on the Earth's surface. It consists in the development of a mathematical apparatus for constructing a correlation model in a spherical coordinate system. For the formation of the model were adopted simplifications, which consist in taking

the form of the Earth in the form of a ball, the satellite's orbit is stationary in the form of a circle over a certain latitude of the Earth. To assess the adequacy of the model an analysis of its results over the elementary types of signals presented in the form of sinusoidal.

## II.TWO-DIMENSIONAL PROCESSES IN CARTESIAN AND SPHERICAL COORDINATE SYSTEMS

The definition of correlation in a spherical coordinate system, by the way, as in the Cartesian or D-denier coordinate system, requires a clear definition whether it is necessary to make correlations of figures or correlation of processes.

In the first case, under the correlation of the figures it is necessary to understand the correlation of the coordinates of the points belonging to a bulk figure with the corresponding coordinates of the points of the standard figure. In this case, it is sufficient to consider the functions describing the studied and standard figures given in parametric or normal form, while the functions depend on time, so the time coordinate is not taken into account. This means that spatial figures remain unchanged in time, therefore, it is sufficient to determine the correlation of only spatial coordinates, in the Cartesian coordinate system or, - in a spherical coordinate system. Then the investigated and standard objects are described in the Cartesian coordinate system by the equations in the normal form:

$$\begin{aligned} z &= f(x, y), \\ z_e &= f_e(x, y). \end{aligned}$$

In a spherical coordinate system:

$$\begin{aligned} r &= \varphi(\Theta, \lambda) \\ r_e &= \varphi_e(\Theta, \lambda) \end{aligned}$$

As a rule, objects describing closed figures in the corresponding space are described, or they specify certain finite surfaces, although they can also describe infinite surfaces that are investigated on a finite spatial range.

One of the least complicated figures in the spherical coordinate system is the sphere with the center at the origin point, which is described by the known equation

$$r(\lambda, \Theta) = R.$$

The sphere in the Cartesian coordinates is described by a more complex expression:

$$x^2 + y^2 + z^2 = R,$$

where  $R$  is the radius of the sphere in the polar and Cartesian coordinate systems.

It is obvious that manipulation with such figure, including the calculation of correlation estimates, is easier to carry out in a spherical coordinate system.

There is a number of figures that are described simply in a spherical system, for example, almost all rotation functions: ellipsoids, paraboloids, hyperboloids, and others.

In the second case when processes are being investigated, while constructing both processes itself and their correlation functions, it is necessary to take into account the time coordinate, since at different times the process may have different spatial coordinates. Therefore, processes need to be investigated in the spatial-temporal coordinate system. Accordingly, it is convenient to represent such processes in a parametric form:

$$\begin{cases} x(t) = f_1(t), \\ y(t) = f_2(t), \\ z(t) = f_3(t). \end{cases} \quad (1)$$

In a spherical coordinate system, the process given parametrically, respectively, is determined by a system of equations:

$$\begin{cases} \lambda(t) = \varphi_1(t), \\ \Theta(t) = \varphi_2(t), \\ r(t) = \varphi_3(t). \end{cases} \quad (2)$$

Then the spherical coordinates in the parametric form have the following form:

$$\begin{aligned} r &= \sqrt{(f_1(t))^2 + (f_2(t))^2 + (f_3(t))^2} \\ \Theta &= \arcctg \frac{f_3(t)}{\sqrt{f_1(t)^2 + f_2(t)^2}} \\ \varphi &= \arcctg \left( \frac{f_2(t)}{f_1(t)} \right) \end{aligned} \quad (3)$$

From (1), (2) and (3) it can be seen that the complexity of expressions will depend on the functions.

We consider a function that describes the sphere in spherical coordinates. This means that, depending on the moment of time, the value of the function will correspond to a certain point on the sphere. This function describes processes in astronomy, in the description of orbits of planets, satellites and other cosmic bodies, the position of objects in radar, guidance systems and a number of other cases. In this case, the expressions will have the simplest form when choosing a coordinate system with the beginning in the scope of the sphere. Such function is given in the polar coordinate system by the equation, where is the radius of the sphere. (fig. 1). When transitioning to the Cartesian coordinate system we use the well-known expressions:

$$\begin{aligned} x &= R \sin \Theta \cos \lambda, \\ y &= R \sin \Theta \sin \lambda, \\ z &= R \cos \Theta. \end{aligned} \quad (4)$$

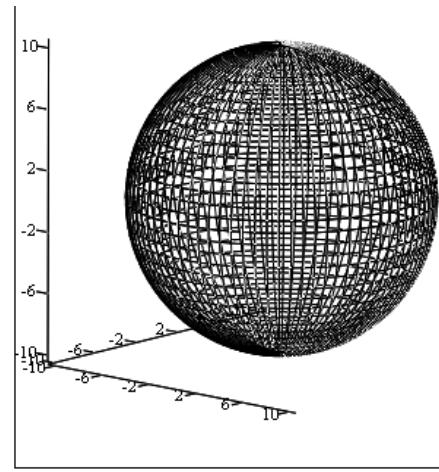


Fig. 1. Graph of the function  $r = R$  given in the polar coordinate system.

Another example of a function that describes a number of processes is:  $r = R$ ,  $\Theta = \alpha$ , where  $R$  – radius,  $\alpha$  – coordinate angle of latitude.

This function describes the circle (fig. 2) and in practice, it can determine the satellite's orbit, which moves over a certain parallel of the Earth. If the center of the Earth is taken as the coordinates' center, then  $R$ - determines the height of the orbit above the surface of the Earth, and the angle of coordinate of latitude is the parallel along which the satellite will move. If,  $\alpha = 0$ , then the orbit will pass over the equator. If  $\Theta = \alpha$  while  $0 < \alpha < \pi/2$  the orbit is passing over a certain parallel in the North hemisphere. If  $\Theta = \alpha$  while  $-\pi/2 < \alpha < 0$  the orbit is passing over a certain parallel in the South hemisphere. If  $\alpha = \pi/2$  or  $\alpha = -\pi/2$ , then the satellite will "hang" over the north or south pole respectively (fig. 2-3).

If the process is periodic with a period  $T$ , then you can determine the periodicity of the coordinates  $\lambda$  and  $\Theta$ , having their own periods of repetition:  $\lambda$  –  $2\pi$  - from 0 to  $2\pi$ ;  $\Theta$  - also  $2\pi$ , from  $-\pi$  to  $\pi$ . Consequently, the coordinate period  $\lambda$  and  $\Theta$  finally will be determined from the functions  $\varphi_1(t)$  for the coordinate  $\lambda$  and the function  $\varphi_2(t)$  - for the coordinate  $\Theta$ . It is obvious that when changing the coordinates of time  $t$  on the segment  $[0, T]$ ,  $[T, T+T]$  etc., respectively, coordinate  $\lambda$  and  $\Theta$  should change for the period  $[0, 2\pi]$ ,  $[2\pi, 4\pi]$ , and  $[-\pi, \pi]$ ,  $[\pi, 3\pi]$ , etc.

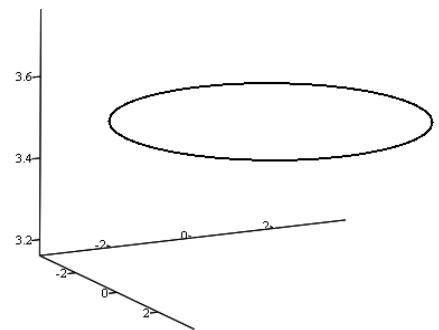


Fig. 2. Schedule of function  $r = R$ ,  $\Theta = \alpha$

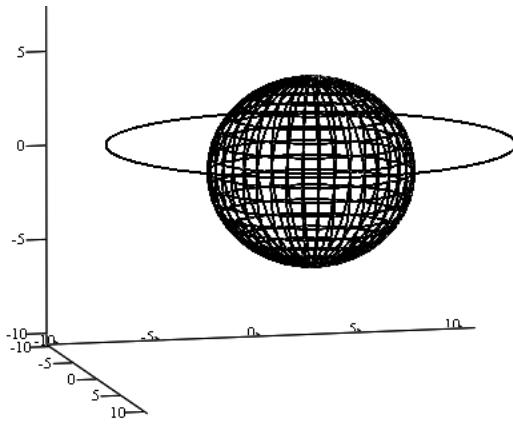


Fig. 3. Graph of an example of a satellite's motion in orbit over a given parallel of the Earth.

If it is difficult or impossible to describe the reference objects or the reference objects functionally, then they are given by an array of coordinate triples,  $x, y, z$  – in the Cartesian coordinate system or  $r, \Theta, \lambda$  – in the spherical coordinate system of the discrete points of the figures and approximated by certain methods.

### III. THEORETICAL FOUNDATIONS OF CONSTRUCTING CORRELATION MODELS IN A SPHERICAL COORDINATE SYSTEM

Having considered the peculiarities of constructing functions in a spherical coordinate system, we turn to the study of the construction of correlation models in a spherical coordinate system and the representation of their graphs in this coordinate system.

Consequently, for the construction of a correlation model of discrete spatial figures [6], the expression for the Cartesian coordinate system will have the form:

$$K_{Re}(p, q) = \frac{1}{m} \cdot \frac{1}{n} \cdot \sum_{i=0}^{N-P-1} \sum_{j=0}^{N-Q-1} f(x_i, y_j) \cdot fe(xe_{i+p}, ye_{j+q}) \quad (5)$$

where  $i, j$  - coordinate indices  $x, y$  working sample points;

$n, m$  – the volumes of working samples according to coordinates  $x, y$ ;

$p, q$  – indices of displacement by the corresponding coordinates,  $x, y$ ;

$P, Q$  – the maximum value of the displacement by the corresponding coordinates  $x, y$ ;

$N$  – the total number of discrete points of an object.

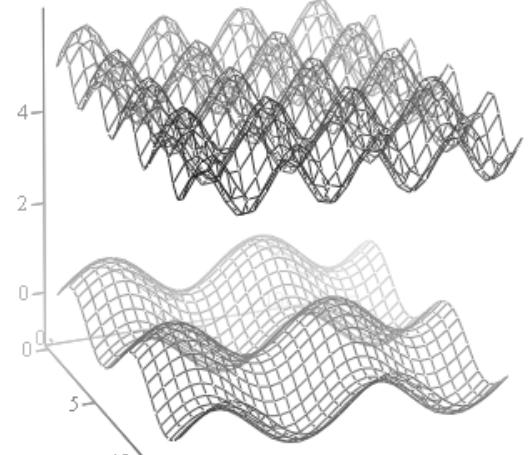
We construct graphs and the corresponding function of correlation (fig. 4), according to (5) for the functions:

$$z1(x, y) = \frac{\sin x + \sin y}{2} \quad \text{and}$$

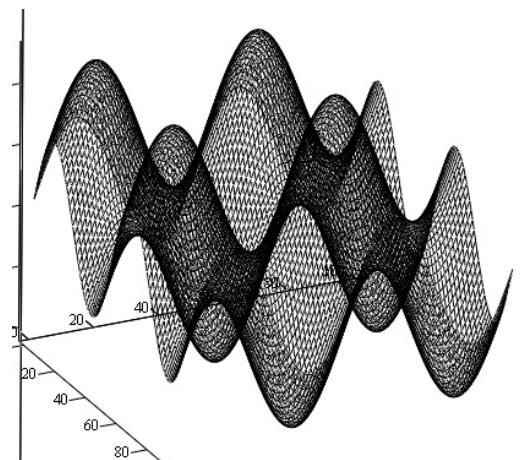
$$z2(x, y) = 5 + \frac{\sin 2x + \sin 2y}{2}.$$

A similar expression for determining the correlation is obtained for a spherical coordinate system:

$$K_{Re}(p, q) = \frac{1}{m} \cdot \frac{1}{n} \sum_{i=0}^{N-P-1} \sum_{j=0}^{N-Q-1} \varphi(\lambda_i, \Theta_j) \cdot \varphi_e(\lambda_{i+p}, \Theta_{j+q}) \quad (6).$$



a)



b)

Fig. 4. a) - Charts functions that correlate  $z1(x, y) = \frac{\sin x + \sin y}{2}$  and  $z2(x, y) = 5 + \frac{\sin 2x + \sin 2y}{2}$ ;

b) - Charts their covariance function.

To represent correlation models in a spherical coordinate system we use formulas (3) to translate Cartesian coordinates into spherical ones.

Most often, correlation models are built for time processes. That is  $F(x, y, z)$ , the function is not a spatial coordinate  $x, y, z$ , but a time-dependent  $t$  function  $f(t)$ . In this case, processes in Cartesian coordinates, are set, as a rule, parametrically (1).

Then spatial coordinates are determined at each time point. In this perspective, it is expedient to consider the time-spatial coordinate system. As already noted, there is a certain category of processes that are much more convenient to represent in a spherical coordinate system. In addition, the analysis of such processes, including correlation, is also

simpler and more convenient in this coordinate system, since mathematical expressions are less complex.

Consider a function that describes the sphere in spherical coordinates. This means that, depending on the moment of time, the value of the function will correspond to a certain point on the sphere. This function describes the processes in astronomy, in the description of orbits of planets, satellites and other cosmic bodies, the position of objects in radar, guidance systems and in a number of other cases. In this case, the expressions will have the simplest form when choosing a coordinate system with the beginning in the center of the sphere. Such functions are given in the polar coordinate system by the equation  $r = R = \text{const}$ , where  $R$  – is the radius of the sphere (Fig. 5).

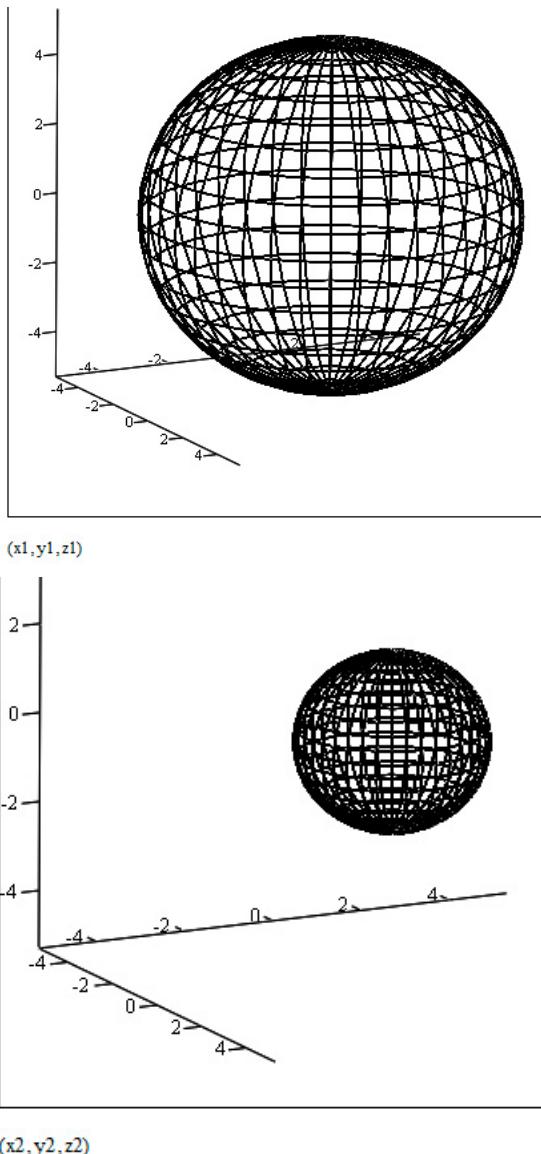


Fig. 5. Charts of functions of the sphere with radii  $R_1 = 5$  and  $R_2 = 2$ .

Using the expression of the covariance function (6), for which it is not necessary to center the values of the random process, unlike the correlation function, we will construct a covariance model, the graph of which is presented in fig. 6.

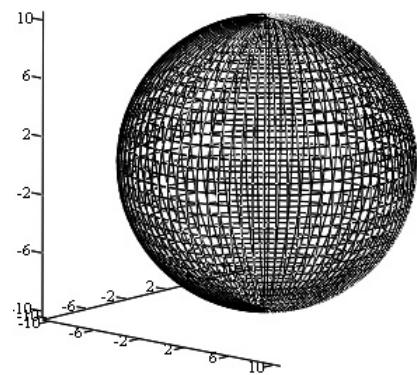


Fig. 6. Representation of a covariance function in a polar coordinate system.

As expected, the covariance function looks like a sphere with a radius  $R = R_1 \cdot R_2 = 2 \cdot 5 = 10$ .

Correlation functions in the polar coordinate system require further research and are an effective research tool.

## VI. CONCLUTION

The method of constructing correlation models in a spherical coordinate system proposed in the article is a very useful tool for the investigation of two-dimensional stochastic and deterministic signals in radar, astronomy, radio engineering, and other spheres. A number of phenomena and processes in a spherical coordinate system are described by simpler mathematical equations, which simplifies their further analysis. The results of the correlation analysis reflected in the spherical coordinate system are often more visible and understandable.

## REFERENCES

- [1] C.W. Gardiner. Handbook of Stochastic Methods: for Physics, Chemistry and the Natural Sciences (Springer Series in Synergetics)/ Hardcover. 2009 XVIII - 447 p.
- [2] Botov M.I. Fundamentals of the theory of radar systems and complexes: Textbook. / M. I. Botov, V. A. Vyakhirev; under the Society. Ed. M.I. Botova. - Krasnoyarsk: Sib. feder. Univ., 2013. - 530 p.
- [3] Shikhovtsev I.V., Yakubov V.P. Statistical radiophysics. Course of lectures / Novosib. state. un-t. Novosibirsk, 2011. – 157 p.
- [4] Babushkin AK, Zakharov NP, Turchaninov AV, Korolev AF Localization of sources of radio emission in media with multipath propagation of radio waves on the basis of correlation measurements // Compilation of reports of the Third All-Russian Conference "Radar and Radiocommunication". M.: Izdatelstvo IRE im. VA Kotelnikova, 2009. PP. 616-619.
- [5] N Shkljarenko, A Segin, J Nikolajchuk, N Terenteva Modern Problems of Radio Engineering, Telecommunications and Computer Science, 2004. Proceedings of the International Conference. PP. 411–412
- [6] Nykolaychuk Y., Segin A. Information source models and methods of there building. Methods and equipment of quality valuation. Ivano-Frankivsk, 1998, № 2, PP. 80 – 84.

# Software for Tolerance Design of Electronic Devices

Galina Shilo, Natalia Furmanova, Tetyana Kulyaba-Kharitonova

Department of Information Technologies of Electronic Devices, Zaporizhzhia National Technical University, UKRAINE,  
Zaporizhzhia, Zhukovskogo str., 64, E-mail: [shilo.gn@gmail.com](mailto:shilo.gn@gmail.com), [nfurmanova@gmail.com](mailto:nfurmanova@gmail.com), [tikylyaba@gmail.com](mailto:tikylyaba@gmail.com)

**Abstract:** The algorithms of tolerance design which are based on the geometric representation of the tolerable domain are proposed. The shape of the domain is defined by law of distributions and optimization criterion. Software for tolerance design is developed. The software can be used as a separate system for mathematical modelling of electronic devices and can be integrated as a component in different electronic design automation systems or computer-aided design systems, for example in SPICE or HFSS.

**Keywords** – tolerable domain, tolerance design, optimization criterion, electronic devices, computer-aided design.

## I. INTRODUCTION

The assignment of tolerances is one of the most important stages in the design of various devices, because standard deviations values of the elements parameters set the accuracy of the output functions and affect the equipment cost. The problem of ensuring the accuracy of output characteristics and parameters of the electronic equipment is solved at the stage of circuit design by tolerance design. Two aspects are to be considered. The deviations of the output functions are to be obtained if the deviations of components parameters are given (tolerance analysis). The deviations of parameters are to be defined if the boundary deviations of the output functions are given (tolerance synthesis).

In modern systems of computer-aided design of electronic devices the procedures of the analysis of tolerances are implemented. These procedures enable to calculate the relative sensitivity of the characteristics of the circuit to change the parameters of the selected element, to perform the Monte Carlo statistical analysis, the simulation of the worst case, to research the circuits taking into account the technological parameters spread of the electronic elements, and temperature dependences. Unfortunately, these systems do not include tools for solving the problems of the tolerances synthesis.

The purpose of the work is to develop software for the tolerance design of electronic devices and integrate it with EDA or ECAD systems.

## II. ALGORITHMS OF THE TOLERANCE DESIGN FOR THE ELECTRONIC EQUIPMENT

Knowledge in the field of tolerance design has been accumulated for many years [1-4]. During this time, it has been established that the value of the tolerances assignment greatly depends on the law of distribution of parameters, which is formed in the process of manufacturing components for different devices. It has led to the widespread using the method of moments, statistical tests of Monte Carlo in tolerances assignment [5].

However, to ensure sufficient accuracy in the synthesis of tolerances, the Monte Carlo method requires a large number of tests. The method of moments is used only in the normal law of the parameters distribution and does not provide sufficient accuracy due to the limited number of parameters of the distribution law (mathematical expectation and dispersion).

The accuracy of the procedure for assigning tolerances was increased when the method of tangent [6] was introduced. In this method, standard deviations of the parameters are formed at the tangent point of tolerance domain and domain of operational capability of device. The interval models of output functions are also used. This approach is successfully used in the procedures for assigning tolerances with a uniform [7], normal law of the parameters distribution [8] or if such a distribution law is given by statistical series [9]. There is a need to implement these approaches in modern computer-aided design systems.

When dealing with issues, the peculiarities of the manufacturing and operation of electronic equipment should be taken into account. It is necessary to provide the possibility to consider the law of distribution of parameters, the correlation between the parameters of the components, and possible compensation of the external factors.

The tolerance synthesis is the inverse problem and therefore it is inaccurate and resolved ambiguously. Thereby, it is necessary to apply methods of optimization according to different criteria for solving this problem. In this case it is convenient to use different algorithms for each design strategy: strategy for equal tolerances ( $\delta$ -strategy), maximum volume tolerance domain ( $V$ - strategy), the minimum cost ( $P$ -strategy), optimal price / quality ratio ( $P/V$ - strategy). Thus, for the software implementation, the following algorithms were developed:

- algorithms for the synthesis of interval tolerances using 4 design strategies;
- algorithms for the synthesis of tolerances on component parameters, taking into account the normal distribution law for 4 design strategies;
- algorithms for the synthesis of tolerances on component parameters taking into account the distribution law given by statistical series for 4 design strategies;
- algorithms for the synthesis of tolerances on the parameters of components, taking into account the correlation between the parameters;
- algorithms for the synthesis of tolerances, taking into account the coefficients of external influences, which are given in the form of interval structures.

The algorithm that enables to assign the same symmetric interval tolerances to the component parameters is following:

**Step 1.** The coefficients of the model of the output function are determined at the point of the parameters nominal values and the initial values of the parameters deviations of the elements are assigned:

$$\underline{\delta}_i^{(0)} = \begin{cases} \frac{\underline{\delta}_y}{n}, & a_{ri} > 0 \\ -\frac{\underline{\delta}_y}{n} & \text{otherwise} \end{cases}, \quad (1)$$

where  $\underline{\delta}_i$  is the relative deviation of component parameters;  $\underline{\delta}_y = (\underline{y} - y_r) / y_r$  is the lower standard deviation of the output function;  $y_r$  is the rating value of the output function;  $\underline{y}$  is the lower value of output function.

**Step 2.** The initial coordinates of the point of tangency between the top of the tolerance domain and the domain boundaries of working operation are determined:

$$\underline{x}_i^{(0)} = x_{ri}(1 + \underline{\delta}_i^{(0)}). \quad (2)$$

**Step 3.** The coefficients of the model of the output function are determined at the point of tangency between the top of the tolerance domain and the boundaries of the capability region. Standard deviations are assigned by the formula

$$\delta = b_w / \sum_{i=1}^n |\bar{a}_i + \underline{a}_i| x_{ri}, \quad (3)$$

where  $b_w = \bar{b} - \underline{b}$ ;  $x_{ri}$  are the rating values of components parameters;  $a_i$  are the model coefficients of the output function at the point of the rating values of parameters.

**Step 4.** The coordinates of the point of tangency between the top of the tolerance domain and the boundaries of the area of working operation are determined:

$$\underline{x}_i^{(k)} = x_{ri}\left(1 + \underline{\delta}_i^{(k)}\right), \quad (4)$$

where  $\underline{\delta}_i^{(k)}$  is the standard deviation of parameters on  $k$  iteration.

**Step 5.** The value of the output function is determined at the tangency point of the top of the tolerance domain and the boundaries of the operational capability region. The completion condition of the algorithm is checked:

$$\left| \frac{y^{(k)} - \underline{y}}{\underline{y}} \right| \leq \varepsilon, \quad (5)$$

where  $\varepsilon$  is accuracy of calculations.

When the condition is satisfied, the algorithm is over, otherwise, step 3 follows.

The algorithms for V-, P- and P/V- strategies include the same steps but deviations of parameters are defined by formulae given in [7]-[10].

In order to take into account external influences on electronic devices, interval structures are used in tolerance synthesis [11]. It allows to store information about a range of changes in external factors and to estimate compensation for their effect. The algorithm for taking into account external

factors is based on the mapping method and has the following form:

**Step 1.**  $\underline{x}_{di}^{(0)}$  boundary values of the parameters of the elements are determined based on  $\underline{y}_e$  boundary operating value of the output function and  $x_{ri}$  rating values of elements parameters. The algorithm for assigning interval tolerances with a given optimization criterion is used.

**Step 2.** In the vicinity of  $\underline{B}_d^{(0)}$  boundary point  $\underline{d}_i$  relative changes in the elements parameters are determined regarding the most unfavorable combination of external factors.

**Step 3.** The coordinates of  $\underline{x}_{di}^{(0)}$  boundary points are mapped onto normal environmental conditions, and the nominal boundary value of the output function is determined in the first approximation:

$$\underline{x}_{ri}^{(0)} = \underline{x}_{di}^{(0)} / \underline{d}_i; \quad \underline{y}_r^{(1)} = y(\underline{X}_r^{(0)}), \quad (6)$$

where  $\underline{X}_r^{(0)} = \{\underline{x}_{r1}^{(0)}, \dots, \underline{x}_{rn}^{(0)}\}$  is the set of coordinates of  $\underline{B}_r^{(0)}$  boundary point.

**Step 4.** The possibility of implementing the algorithm for a given boundary operational value of the output function and given coefficients of external influences is tested. To do this, the condition is checked:

$$\underline{y}_r^{(1)} < y_r, \quad (7)$$

where  $y_r = y(X_r)$  is the rating value of the output function.

If the condition is satisfied, the algorithm ends and the message about the impossibility of implementation is displayed.

**Step 5.** Nominal interval tolerances are assigned at the boundary values of  $\underline{y}_r^{(k)}$  output function and  $\underline{x}_{ri}^{(k)}$  rating boundary values of the elements parameters are determined. The algorithm for assigning interval tolerances is used.

**Step 6.** The boundary values of the elements parameters and the output function are determined by the effect of external factors:

$$\underline{x}_{di}^{(k)} = \underline{x}_{ri}^{(k)} \underline{d}_i; \quad \underline{y}_d^{(k)} = y(\underline{X}_d^{(k)}). \quad (8)$$

**Step 7.** The relative change in the boundary value of the output function is determined and its boundary nominal value is specified:

$$\begin{aligned} \underline{d}_y^{(k)} &= \underline{y}_d^{(k)} / \underline{y}_r^{(k)}; \\ \underline{y}_r^{(k)} &= \underline{y}_r^{(k-1)} - (\underline{y}_d^{(k)} - \underline{y}_e) / \underline{d}_y^{(k)}. \end{aligned} \quad (9)$$

**Step 8.** The condition for the completion of the algorithm is checked:

$$\left| \frac{\underline{y}_d^{(k)} - \underline{y}_e}{\underline{y}_e} \right| \leq \varepsilon, \quad (10)$$

where  $\varepsilon$  is the accuracy of calculation.

If the condition is satisfied, the algorithm ends, otherwise, go to step 5.

These algorithms are used to develop the software for tolerance design.

### III. TOLERANCE DESIGN SOFTWARE

Tolerance design software are allowed to analyze and synthesize deviations of geometric and electric parameters of electronic devices and select components. The main features are

- analysis of deviations of output function by given standard deviation of parameters of components;
- synthesis of equal deviations of the parameters;
- synthesis of deviations of the parameters for case of maximal volume of tolerance domain;
- selection of components by optimal price/quality ratio.

Software consists of the modules. They are the forming of the models of output characteristics, tolerance analysis, tolerance synthesis, component selection, report generation.

The structure of the software is shown in Fig. 1.

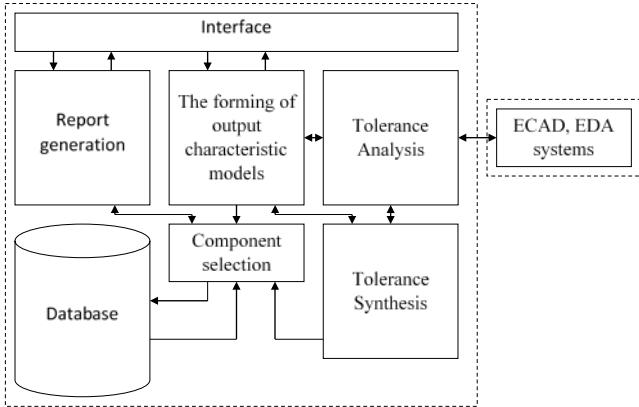


Fig. 1. The structure of the software for the tolerance design

The Database is used for component selection. Tolerance synthesis module includes

- selection of optimality criterion;
- procedures for synthesis of the tolerances of electronic device parameters for the uniform law of parameter distribution, the normal law, and distribution given by statistical series;
- procedures for synthesis of the tolerances taking account correlation between parameters;
- procedures for synthesis of the operating tolerances;

Input data are deviations of output function, the law of parameter distribution, rating values of parameters, coefficients of external factors, accuracy of calculation. As a result, the tolerances of parameters are defined. In the software it is necessary to choose the mode of output function forming in the program system. There are the symbolic form mode and the mode of forming models by external ECAD system.

The module for the formation of output characteristics enables to describe the mathematical models in a symbolic form. The formulae are created in two modes: command mode and visual ones. MathML library is used for visualization.

An example of the application of tolerance design software for synthesize the deviations of the parameters of the power pulse generator is given below. Pulse duration is defined by

$$e^{\frac{\tau}{R_3 C_1}} = 1 - e^{\frac{\tau}{R_2 C_2}}.$$

The rating values of parameters are the following  $R_2 = 110 \text{ kOhm}$ ;  $R_3 = 110 \text{ kOhm}$ ;  $C_1 = 0,47 \text{ mF}$ ;  $C_2 = 0,47 \text{ mF}$ . The rating value of pulse duration is  $\tau_r = 35,84 \text{ ms}$ .

If the deviation of pulse duration is  $\Delta\tau = \pm 1 \text{ ms}$ , then standard deviations for the elements parameters are  $\delta_R = \delta_C = \pm 1,43 \%$ .

If the deviation of the pulse duration is  $\Delta\tau = \pm 2 \text{ ms}$ , then standard deviations for the parameters are  $\delta_R = \delta_C = \pm 2,86 \%$ . For the boundary values of the delay time, the values of tolerances on the parameters of the scheme components are given in Table 1 and Table 2.

TABLE 1. BOUNDARY VALUES OF DELAY TIME FOR THE DEVIATION OF  $\Delta\tau = \pm 1 \text{ MS}$  PULSE DURATION.

| Delay time         | Boundary deviations of parameters, % |               |               |               | Deviations of delay time, ms |
|--------------------|--------------------------------------|---------------|---------------|---------------|------------------------------|
|                    | $\delta_{C1}$                        | $\delta_{R3}$ | $\delta_{R2}$ | $\delta_{C2}$ |                              |
| $\bar{\tau}$       | 1,43                                 | 1,43          | 1,43          | 1,43          | 1,031                        |
| $\underline{\tau}$ | -1,43                                | -1,43         | -1,43         | -1,43         | -1,017                       |

TABLE 2. BOUNDARY VALUES OF DELAY TIME FOR THE DEVIATION OF  $\Delta\tau = \pm 2 \text{ MS}$  PULSE DURATION.

| Delay time         | Boundary deviations of parameters, % |               |               |               | Deviations of delay time, ms |
|--------------------|--------------------------------------|---------------|---------------|---------------|------------------------------|
|                    | $\delta_{C1}$                        | $\delta_{R3}$ | $\delta_{R2}$ | $\delta_{C2}$ |                              |
| $\bar{\tau}$       | 2,86                                 | 2,86          | 2,86          | 2,86          | 2,077                        |
| $\underline{\tau}$ | -2,86                                | -2,86         | -2,86         | -2,86         | -2,019                       |

If the component parameters are distributed according to the normal distribution law, then for the case  $\Delta\tau = \pm 1 \text{ ms}$ ; tolerances are  $\delta_R = \delta_C = \pm 2,74 \%$ . Delay time values are  $\bar{\tau} = 1,031 \text{ ms}$ ;  $\underline{\tau} = -1,017 \text{ ms}$ . If  $\Delta\tau = \pm 2 \text{ ms}$ , then  $\delta_R = \delta_C = \pm 5,47 \%$  and boundary values of delay time are  $\bar{\tau} = 2,077 \text{ ms}$ ;  $\underline{\tau} = -2,019 \text{ ms}$ .

Integration of the developed automated system with modern CAD is possible in three ways:

- by developing the interface software module, which uses specialized macros to calculate the output characteristics of the radio electronic device in CAD environment;
- by developing an internal application in CAD, which implements the calculation of output characteristics of radio-electronic devices using API functions;
- by inputting the value of the output characteristic, calculated using a special CAD system in a dialogue mode.

The first integration method is implemented in the ANSYS HFSS system for the low frequency coaxial filter. The example of program interface is shown in Fig. 2. LPF is designed to reduce the level of side-effects in the spectrum of probe signals and suppression of off-band radiation in the spectrum of output signals of amplifier modules of a distributed transmitting device in the antenna.

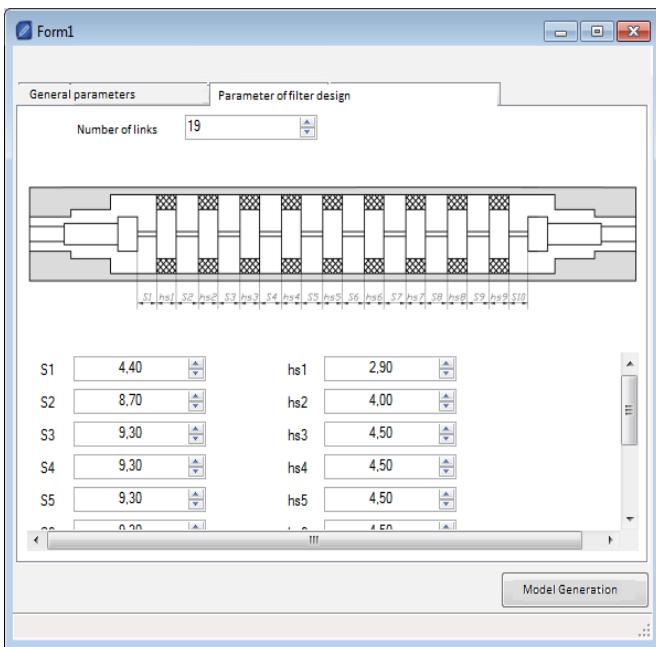


Fig. 2. Program interface of tolerance design software

For an example of the applying of the tolerance assignment system, a coaxial low pass filter (LPF) with following parameters was designed:

- number of links -19, cut-off frequency of the LPF ~ 3300 MHz;
- bandwidth from 2800 to 3100 MHz;
- maximum losses in the bandwidth are not more than 0,3 dB;
- the static wave constant at the voltage in the bandwidth  $K \leq 1,2$ ;
- effective attenuation in the barrier band (for the 2nd and 3rd harmonics band) - not less than 60 dB: from 5600 to 6200 MHz; from 8400 to 9300 MHz.

As result the three-dimensional model of filter is generated. This model is shown in Fig.3.

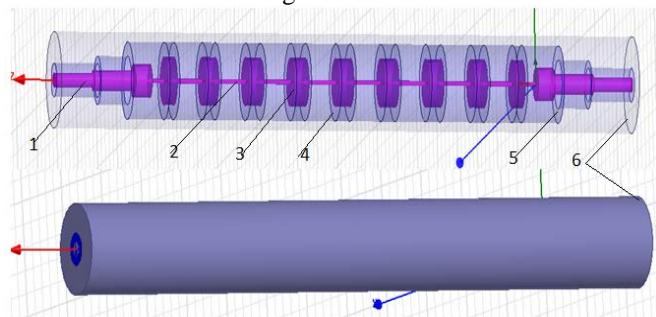


Fig.3. The three-dimensional model of coaxial low pass filter.

Analysis of the influence of the parameters tolerances of the step transition to the standing wave coefficient by the interval methods was carried out.

#### IV. CONCLUSION

The software of tolerance design has been developed. It provides an opportunity to analyze and synthesize tolerances if the tolerance domain is represented by various geometric objects (boxes, ellipsoids, and their combination). Individual procedures are implemented to assign tolerances of parameters taking into account the correlation between the parameters, the effect of external factors during the operation stage. The possibility to calculate tolerances for different design strategies is provided. These strategies ensure either the best quality or minimum cost, or taking into account the features of the technological process (strategy of equal tolerances).

#### REFERENCES

- [1] B. Mandziy, B. Volochiy, L. Ozirkovskyy "Program models for interactive design of fault-tolerant system with mixed structural redundancy taking account maintenance service strategies," *Computing*, Vol. 7, Issue 1, 2008, pp.161-163.
- [2] R. Spence and R.S. Soin, *Tolerance Design of Electronic Circuits*, World Scientific, 2002, 232 p.
- [3] J.B. Grimbleby, *Computer-aided analysis and design of electronic networks*, London: Pitman, 1990, 289 p.
- [4] L. Kolev, "Worst-case tolerance analysis of linear DC and AC electric circuits", *IEEE Trans.Circuits Systems*, vol. 47, 2002, pp. 1 – 9.
- [5] J. Gajda and T. Sidor, "Using Monte Carlo Analysis for Practical Investigation of Sensitivity of Electronic Converters in Respect to Component Tolerances," *Electrical And Electronic Engineering* Vol. 2 (5), 2012, pp. 297-302.
- [6] G.Shilo Geometric Methods of Tolerances Setting //Journal of Automation and information sciences –2007. – Vol. 39. – № 3. – P. 51-60.
- [7] G.M.Shilo, "Strategies for Assigning Interval Tolerances" *Cybernetics and Systems Analysis*, vol. 51, issue 4, 2015, pp. 657-666.
- [8] G.N.Shilo, "Normal tolerance assigning by given price characteristics of radio components" *Radioelectronics and Communications Systems* Volume 55, Issue 3, March 2012, Pages 140-148
- [9] G.Shilo, "Assigning Tolerances by Method of Smoothed Vertices," *Journal of Automation and Information Sciences*, Begell House, Inc, vol. 45 , issue 10, 2013, pp.36-48.
- [10] G.Shilo, N.Furmanova "Statistically Oriented Tolerance Design with Correlation between Parameters of Components" *Proceedings of the 9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications* 21-23 September, 2017, Bucharest, Romania. – p.1082 – 1087.
- [11] G.M.Shilo, "Interval structures and their applications" *Computing*, vol 6, issue 1,2007, pp. 80-87.

# Mathematical Model of Balanced Layout Problem Using Combinatorial Configurations

Igor Grebennik<sup>1</sup>, Tatiana Romanova<sup>2</sup>, Inna Urniaieva<sup>1</sup>, Sergey Shekhovtsov<sup>1</sup>

1. Department of Computer Science, Kharkiv National University of Radioelectronics, UKRAINE, Kharkiv, 14 Nauki ave., email: [igorgrebennik@gmail.com](mailto:igorgrebennik@gmail.com)

2. Department of Mathematical Modeling and Optimal Design, Institute for Mechanical Engineering Problems of the National Academy of Sciences of Ukraine, UKRAINE, Kharkiv, 2/10 Pozharskogo str., email: [tarom7@yahoo.com](mailto:tarom7@yahoo.com)

**Abstract:** The optimization of the balanced layout of a set of 3D-objects in a container is considered. We define combinatorial configurations describing the combinatorial structure of the problem. A mathematical model of the problem is presented. The model takes into account the placement constraints, the mechanical characteristics and the combinatorial features of the problem.

**Keywords:** Balanced Layout, 3D-objects, Combinatorial Configurations, Phi-function Technique.

## I. INTRODUCTION

Balanced layout problems belong to the class of NP-hard placement problems [1] and are the subject of the study of computational geometry, and the methods for their solution are a new branch of the theory of operations research [2, 3]. The essence of the problem lies in the search for the optimal placement of a given set of 3D-objects in a container, taking into account so called, behavior constraints, which ensure the balance of the system under consideration [4], [5].

## II. PROBLEM FORMULATION

Let  $\Omega$  be a container of height  $H$  that can take the form of a cuboid, cylinder, paraboloid of rotation, or truncated cone. The container  $\Omega$  is defined in the global coordinate system  $Oxyz$ , where  $Oz$  is the longitudinal axis of symmetry. We assume that container  $\Omega$  is divided by horizontal racks into sub-containers  $\Omega^j$ ,  $j \in J_m = \{1, \dots, m\}$ . We denote distances between racks  $S_j$

and  $S_{j+1}$  by  $t_j$ ,  $j \in J_m$ ,  $\sum_{j=1}^m t_j = H$ . The center of the

lower base of container  $\Omega$  is located in the origin of the global coordinate system  $Oxyz$ .

Let  $A = \{\mathbb{T}_i, i \in J_n\}$  be a set of homogeneous 3D-objects given by their metrical characteristics. Each object  $\mathbb{T}_i$  of height  $h_i$  and mass  $m_i$ , is defined in its local coordinate system  $O_i x_i y_i z_i$ ,  $i \in J_n$ . The location of object  $\mathbb{T}_i$  inside container  $\Omega$  is defined by vector  $u_i = (v_i, z_i, \theta_i)$ , where  $(v_i, z_i)$  is a translation vector in the global coordinate system  $Oxyz$ ,  $\theta_i$  is a rotation angle of object  $\mathbb{T}_i$  in the plane  $O_i x_i y_i$ , where  $v_i = (x_i, y_i)$ , and the value of  $z_i$ ,

$i \in J_n$ , is uniquely defined by sub-container  $\Omega^j$ ,  $j \in J_m$ , in which object  $\mathbb{T}_i$  will need to be placed.

In contrast to the BLP problems, where a priori the requirement for placing objects in specific sub-containers  $\Omega^j$ ,  $j \in J_m$ , is known, in this study the problem of the balanced layout of objects is formulated, which involves generation and selection of a partition of the set  $A$  into non-empty subsets  $A^j$ ,  $j \in J_m$ . Here  $A^j$  is a subset of objects which have to be placed on rack  $S_j$  inside  $\Omega^j$ .

On placement of object  $\mathbb{T}_i$ ,  $i \in J_n$ , inside subcontainer  $\Omega^j$  the following constraints are imposed

$$z_i = \sum_{l=1}^j t_{l-1} + h_i, \quad (1)$$

where  $j \in J_m$ . We consider that  $t_0 = 0$  and  $\forall i \in J_n$  there exists  $j^* \in J_m$ :  $h_i \leq t_{j^*}$ .

Let  $J_n^j \subseteq J_n$  be a set of indexes of objects which are placed in sub-container  $\Omega^j$ ,  $j \in J_m$ ,

$$\bigcup_{j=1}^m J_n^j = J_n, \quad J_n^i \cap J_n^j = \emptyset, \quad i \neq j \in J_m; \quad (2)$$

$k_j = |A^j|$  is the number of objects which are placed in sub-container  $\Omega^j$ ,  $k_j > 0$ ,  $j \in J_m$ ,

$$\sum_{j=1}^m k_j = n. \quad (3)$$

In addition, the following placement constraints have to be taken into account:

$$\text{int } \mathbb{T}_{i_1} \cap \text{int } \mathbb{T}_{i_2} = \emptyset, \quad i_1 < i_2 \in J_n^j, \quad j \in J_m, \quad (4)$$

$$\mathbb{T}_i \subset \Omega^j, \quad i \in J_n^j, \quad j \in J_m, \quad (5)$$

$$h^j \leq t_j, \quad h^j = \max\{h_i^j, i \in J_n^j\}, \quad j \in J_m. \quad (6)$$

We designate a system, formed as a result of the placement of objects  $\mathbb{T}_i$  of the set  $A$  in container  $\Omega$  by  $\Omega_A$ , and a system of coordinates of  $\Omega_A$  by  $O_sXYZ$ , where  $O_s = (x_s(v), y_s(v), z_s(v))$  is the mass center of  $\Omega_A$

$$x_s(v) = \frac{\sum_{i=1}^n m_i x_i}{M}, \quad y_s(v) = \frac{\sum_{i=1}^n m_i y_i}{M}, \quad z_s = \frac{\sum_{i=1}^n m_i z_i}{M}, \quad (7)$$

$M = \sum_{i=1}^n m_i$  is the mass of system  $\Omega_A$  and  $O_s X \parallel O_x$ ,  $O_s Y \parallel O_y$ ,  $O_s Z \parallel O_z$ .

We consider the deviation of the center of mass  $O_s$  of system  $\Omega_A$  from given point  $(x_0, y_0, z_0)$  as an objective function.

*Combinatorial Balanced layout Problem (CBLP).* Define such variant of the partition of the object set  $A$  into nonempty subsets  $A^j$ ,  $j \in J_m$ , and the corresponding placement parameters  $u_i = (v_i, z_i, \theta_i)$  of objects  $T_i$ ,  $i \in J_n$ , taking into account relations (2)–(6), that the objective function will reach its minimum value.

We assume that the problem has at least one feasible solution.

*Note.* Restrictions on the axial and centrifugal moments of the system and allowable distances between objects may also be given.

### III. MATHEMATICAL MODEL

Now we define special combinatorial configurations describing the discrete structure of the CBLP problem. Some basic approaches for mathematical modelling of optimization problems on combinatorial configurations are described in e.g., [6, 7].

The variants of partition of the set  $A$  into non-empty subsets  $A^j$ ,  $j \in J_m$ , are determined by both the number of elements in each subset and the order of the subsets.

Let us consider the sub-containers and the assumed corresponding sets of objects  $A^j$ ,  $j \in J_m$ . Then the tuple of

natural numbers  $(k_1, k_2, \dots, k_m)$ , such that  $\sum_{j=1}^m k_j = n$ ,

determines possible number  $k_j$  of objects in each sub-container  $\Omega^j$ . The number of all such tuples is equal to the number of compositions of the number  $n$  of length  $m$  [8], which is  $|C_{n-1}^{m-1}|$ .

Let us derive in what ways it is possible to decompose  $n$  various objects from a set  $A$  into  $m$  sub-containers  $\Omega^j$ ,  $j \in J_m$ , if in sub-containers there are accordingly  $k_1, k_2, \dots, k_m$  objects, and sets of objects  $A^j$ ,  $j \in J_m$ , inside corresponding sub-containers  $\Omega^j$ ,  $j \in J_m$ , are not ordered.

Without loss of generality, we will distinguish the objects with the same values of metrical characteristics, height  $h_i$  and mass  $m_i$  (for example, consider them to be different in number).

We order the elements of set  $A$ . To each object we assign the number of the sub-container into which it is

expected to be placed. We get a tuple consisting of  $n$  elements that form a permutation with repetitions from  $m$  numbers  $1, 2, \dots, m$ , in which the first element is repeated  $k_1$  times, the second element is repeated  $k_2$  times, ..., the last element is repeated  $k_m$  times.

The total number of permutations is equal to

$$P(n, k_1, k_2, \dots, k_m) = \frac{n!}{k_1! k_2! \dots k_m!}. \quad (8)$$

Then the number of variants of partitions of various objects from set  $A$  to  $m$  sub-containers  $\Omega^j$ , provided that each sub-container contains at least one object and the order of placing objects inside the sub-container is not important, is equal to

$$\sum_{k_1+k_2+\dots+k_m=n} P(n, k_1, k_2, \dots, k_m) = \sum_{k_1+k_2+\dots+k_m=n} \frac{n!}{k_1! k_2! \dots k_m!}$$

Note that the number of summands in the sum is equal to  $N = |C_{n-1}^{m-1}|$ .

To generate subsets  $A^j$ ,  $j \in J_m$ , we introduce a special combinatorial configuration [9].

Rather complex combinatorial configurations can formally be described and generated using tools of construction of compositional  $\kappa$ -images of combinatorial sets ( $\kappa$ -sets) proposed in [10]. A combinatorial set is considered as a set of tuples that constructed from a finite set of arbitrary elements (so-called generating elements) according to certain rules. Permutations, combinations, arrangements, and binary sequences are the examples of classical combinatorial sets.

The basic idea of generation of  $\kappa$ -sets is introduced in [10]. However, the problem of generating  $\kappa$ -sets of more complicated combinatorial structure remains the open problem. Just one of such special cases is studied in [11].

The problem of generating  $\kappa$ -sets is based on special techniques of generating base combinatorial sets. The base sets can be defined as combinatorial sets with the known descriptions: both classical combinatorial sets (permutations, combinations, arrangements, tuples) or non-classical combinatorial sets (permutations of tuples, compositions of permutations, permutations with a prescribed number of cycles, etc.). Generation algorithms for some of base combinatorial sets are presented in, e.g., [12-15].

We denote  $\mathcal{C}_P(n, m)$  the set of compositions of the number  $n$  of length  $m$  (which corresponds to the partition of different objects from set  $A$  to  $m$  sub-containers  $\Omega^j$ ,  $j \in J_m$ , provided that each sub-container contains at least one object and the order of objects inside the sub-container is not important). Wherein,  $|\mathcal{C}_P(n, m)| = N = |C_{n-1}^{m-1}|$ .

Let  $\mathbf{k} = (k_1, \dots, k_m) \in \mathcal{C}_P(n, m)$ ,  $\sum_{j=1}^m k_j = n$ ,  $k_j \geq 1$ ,  $j \in J_m$ .

We introduce a combinatorial set  $\mathcal{Q}(\mathbf{k})$  that is a composition image of combinatorial sets ( $\kappa$ -set)  $\mathcal{C}_P(n, m); C_n^{k_1}, C_{n_1}^{k_2}, C_{n_2}^{k_3}, \dots, C_{n_{m-1}}^{k_m}$ , generated by sets

$I_{n_0}, I_{n_1}, I_{n_2}, \dots, I_{n_{m-1}}$ , where  $n_i = n - k_1 - \dots - k_i$ ,  
 $i \in J_{m-1}$ ,  $I_{n_0} = J_n$ ,

$$I_{n_1} = I_{n_0} \setminus \{j_1^{n_0}, j_2^{n_0}, \dots, j_{k_1}^{n_0}\}, (j_1^{n_0}, j_2^{n_0}, \dots, j_{k_1}^{n_0}) \in C_n^{k_1},$$

$$I_{n_2} = I_{n_1} \setminus \{j_1^{n_1}, j_2^{n_1}, \dots, j_{k_2}^{n_1}\}, (j_1^{n_1}, j_2^{n_1}, \dots, j_{k_2}^{n_1}) \in C_{n_1}^{k_2},$$

...

$$I_{n_{m-1}} = I_{n_{m-2}} \setminus \{j_1^{n_{m-2}}, j_2^{n_{m-2}}, \dots, j_{k_{m-1}}^{n_{m-2}}\},$$

$$(j_1^{n_{m-2}}, j_2^{n_{m-2}}, \dots, j_{k_{m-1}}^{n_{m-2}}) \in C_{n_{m-2}}^{k_{m-1}},$$

$$I_{n_{m-1}} = \{j_1^{n_{m-1}}, j_2^{n_{m-1}}, \dots, j_{k_m}^{n_{m-1}}\},$$

$$(j_1^{n_{m-1}}, j_2^{n_{m-1}}, \dots, j_{k_m}^{n_{m-1}}) \in C_{n_{m-1}}^{k_m}.$$

Note that

$$I_{n_0} \cup I_{n_1} \cup \dots \cup I_{n_{m-1}} = J_n = \{1, 2, \dots, n\},$$

$$I_{n_s} \cap I_{n_t} = \emptyset, s \neq t \in J_{m-1}^0 = \{0, 1, \dots, m-1\}.$$

Each element  $q(\mathbb{k}) \in \mathcal{Q}(\mathbb{k})$  can be described in the form

$$q(\mathbb{k}) = (q_1, \dots, q_{k_1} | q_{k_1+1}, \dots, q_{k_1+k_2} | \dots, | q_{k_1+\dots+k_{m-1}}, \dots, q_{k_{m-1}+k_m}),$$

$$\text{where } (q_1, \dots, q_{k_1}) = (j_1^{n_0}, j_2^{n_0}, \dots, j_{k_1}^{n_0}) \in C_n^{k_1},$$

$$(q_{k_1+1}, \dots, q_{k_1+k_2}) = (j_1^{n_1}, j_2^{n_1}, \dots, j_{k_2}^{n_1}) \in C_{n_1}^{k_2},$$

...

$$(q_{k_1+\dots+k_{m-1}}, \dots, q_{k_{m-1}+k_m}) = (j_1^{n_{m-1}}, j_2^{n_{m-1}}, \dots, j_{k_m}^{n_{m-1}}) \in C_{n_{m-1}}^{k_m}$$

The cardinality of set  $\mathcal{Q}(\mathbb{k})$  is derived by (9).

An element  $q(\mathbb{k})$  of the set  $\mathcal{Q}(\mathbb{k})$  is said to be a *tuple of partition* of the set  $A$  into subsets  $A^j, j \in J_m$ .

Now we define the vector of the basic variables of the problem CBLP:  $u = (v, z, \theta)$ , where  $v = (v_1, \dots, v_n) \in \mathbf{R}^{2n}$ ,  $\theta = (\theta_1, \dots, \theta_n) \in \mathbf{R}^n$ ,  $v_i = (x_i, y_i) \in \mathbf{R}^2$ ,  $x_i, y_i, \theta_i$  are continuous variables,  $z = (z_1, \dots, z_n) \in \mathbf{R}^n$ ,  $z_i, i \in J_n$ , are discrete variables defined by (1).

The values of variables  $z_i, i \in J_n$ , are determined in the order given by elements  $q(\mathbb{k})$  of combinatorial set  $\mathcal{Q}(\mathbb{k})$ :

$$z_{q_i} = \sum_{l=1}^s t_{l-1} + h_{q_i}, \quad (10)$$

where

$$s = \begin{cases} 1, & \text{if } i \leq k_1, \\ 2, & \text{if } k_1 < i \leq k_1 + k_2, \\ \dots \\ m, & \text{if } k_1 + k_2 + \dots + k_{m-1} < i \leq k_1 + k_2 + \dots + k_m, \end{cases}$$

$$i = 1, 2, \dots, n, q_i \in \{1, 2, \dots, n\}, q(\mathbb{k}) \in \mathcal{Q}(\mathbb{k}).$$

Let us formalize placement constraints (4)-(6), using phi-function technique.

We consider two objects  $\mathbb{T}_1$  and  $\mathbb{T}_2$  with variable parameters  $u_1 = (v_1, z_1, \theta_1) \in \mathbf{R}^3$ ,  $u_2 = (v_2, z_2, \theta_2) \in \mathbf{R}^3$ , where  $v_1 = (x_1, y_1)$ ,  $v_2 = (x_2, y_2)$ ,  $x_1, y_1, \theta_1$ ,  $x_2, y_2, \theta_2$  are continuous variables and  $z_1, z_2$  are discrete variables.

By definition [2, 3] a phi-function is a continuous function, therefore we extend the concept to discrete variables  $z_1, z_2$ .

*Definition 1.* Function  $\Upsilon_{12}(u_1, u_2)$  is called a *D-phi-function* of 3D-objects  $\mathbb{T}_1$  and  $\mathbb{T}_2$  if function  $\Upsilon_{12}(v_1, z_1^0, \theta_1, v_2, z_2^0, \theta_2)$  is a *phi-function*  $\Phi_{12}(v_1, z_1^0, \theta_1, v_2, z_2^0, \theta_2)$  of objects  $\mathbb{T}_1$  and  $\mathbb{T}_2$  for fixed values  $z_1 = z_1^0$  and  $z_2 = z_2^0$ .

*Definition 2.* Function  $\Upsilon'_{12}(u_1, u_2, u_{12})$  is called a *quasi D-phi-function* of 3D-objects,  $\mathbb{T}_1$  and  $\mathbb{T}_2$  if function  $\Upsilon'_{12}(v_1, z_1^0, \theta_1, v_2, z_2^0, \theta_2, u_{12})$  is a *quasi-phi-function*  $\Phi'_{12}(v_1, z_1^0, \theta_1, v_2, z_2^0, \theta_2, u_{12})$  of objects  $\mathbb{T}_1$  and  $\mathbb{T}_2$  for fixed values  $z_1 = z_1^0$  and  $z_2 = z_2^0$ .

Here  $u_{12}$  is the vector of auxiliary continuous variables that is used to constructs a quasi *phi-function* of two objects  $\mathbb{T}_1$  and  $\mathbb{T}_2$ .

The placement constraints (4)-(6) are described by the system of inequalities  $\Upsilon_1(u, \tau) \geq 0$ ,  $\Upsilon_2^*(u) \geq 0$ , where the inequality  $\Upsilon_1(u, \tau) \geq 0$  describes the non-overlapping constraints and the inequality  $\Upsilon_2^*(u) \geq 0$  describes the containment constraints

$$\Upsilon_1(u, \tau) = \min \{\Upsilon_1^j(u, \tau), j \in J_m\},$$

$$\Upsilon_1^j(u, \tau) = \min \{\Upsilon_{q_1 q_2}^j(u_{q_1}, u_{q_2}, u_{q_1 q_2}), q_1 < q_2 \in J_n^j\}, \quad (11)$$

$$\tau = (u_{q_1 q_2}, q_1 < q_2 \in J_n^j), \quad \Upsilon_2^*(u) = \min \{\Upsilon_2^{*j}(u), j \in J_m\},$$

$$\Upsilon_2^{*j}(u) = \min \{\Upsilon_{q_i}^*(u_{q_i}), q_i \in J_n^j\}, \quad (12)$$

$\Upsilon_{q_1 q_2}^j(u_{q_1}, u_{q_2}, u_{q_1 q_2})$  is the function that describes non-overlapping of objects  $\mathbb{T}_{q_1}$  and  $\mathbb{T}_{q_2}$ ,  $u_{q_1} = (x_{q_1}, y_{q_1}, z_{q_1}, \theta_{q_1})$ ,  $u_{q_2} = (x_{q_2}, y_{q_2}, z_{q_2}, \theta_{q_2})$ ,

$\Upsilon_{q_i}^*(u_{q_i})$  is the function that describes non-overlapping of objects  $\mathbb{T}_{q_i}$  and  $\Omega^{*j} = \mathbf{R}^3 / \text{int } \Omega^j$ .

Thus, in relations (11), (12) for fixed values  $z_{q_1}$  and  $z_{q_2}$ , we have:  $\Upsilon_{q_1 q_2}^j(u_{q_1}, u_{q_2})$  is a *phi*-function [16]  $\Phi_{q_1 q_2}^{\mathbb{T}\mathbb{T}}(u_{q_1}, u_{q_2})$  for objects  $\mathbb{T}_{q_1}$  and  $\mathbb{T}_{q_2}$  or a quasi-*phi*-function [17]  $\Phi'_{q_1 q_2}^{\mathbb{T}\mathbb{T}}(u_{q_1}, u_{q_2}, u_{q_1 q_2})$  for objects  $\mathbb{T}_{q_1}$  and  $\mathbb{T}_{q_2}$ ;  $\Upsilon_{q_i}^*(u_{q_i})$  is a *phi*-function  $\Phi_{q_i}^{\mathbb{T}\Omega^{*j}}(u_{q_i})$  for objects  $\mathbb{T}_{q_i}$  and  $\Omega^{*j}$ .

If the minimum allowable distances between objects are given, adjusted *phi*-functions (quasi-*phi*-functions) are used for the corresponding pairs of objects [16, 17].

Mathematical model of the CBLP problem can be defined as follows:

$$F(u^*, \tau^*) = \min F(u, \tau) \text{ s.t. } (u, \tau) \in W, \quad (13)$$

$$W = \{(u, \tau) \in \mathbf{R}^\sigma : \Upsilon_1(u, \tau) \geq 0, \Upsilon_2^*(u) \geq 0, \mu(u) \geq 0\}, \quad (14)$$

where

$$F(u) = d = (x_s(v, z) - x_0)^2 + (y_s(v, z) - y_0)^2 + (z_s - z_0)^2$$

$$u = (v, z, \theta), \quad v = (v_1, \dots, v_n), \quad \theta = (\theta_1, \dots, \theta_n),$$

$$v_i = (x_i, y_i), \quad i \in J_n, \quad z = (z_1, \dots, z_n),$$

function  $\Upsilon_1(u, \tau)$  is described by (11) with  $\Xi = \bigcup_{j=1}^m \Xi^j$ ,

$$\Xi^j = \{(q_1, q_2) : q_1 < q_2 \in J_n^j\},$$

$\tau = (\tau_1, \dots, \tau_s) = (u_{q_1 q_2}, q_1 < q_2 \in J_n^j)$  is a vector of auxiliary variables for quasi phi-functions,  $s = |\Xi|$ , function  $\Upsilon_2^*(u)$  is defined by (12), elements of vector  $z$  are given by (10),  $\mu(u) \geq 0$  describes behavior constraints.

CBLP problem can be represented as a mixed integer programming (MIP) problem, using approach with boolean variables.

However, unlike (13)-(14), this approach leads to increasing the number of discrete variables of the model and therefore increases the dimension of the CBLP problem in MIP form.

#### IV. CONCLUSION

We study the problem of the balanced layout of 3D-objects within a container divided by horizontal racks onto sub-containers.

A mathematical model has been constructed that takes into account not only the geometrical and behavior constraints, but also combinatorial features of the problem

associated with the construction of partitions of the set of placed objects into sub-containers.

#### REFERENCES

- [1] B. Chazelle, H. Edelsbrunner, L. Guibas, "The complexity of cutting complexes". *Discrete & Comp. Geom.* 4(2), pp. 139–181, 1989
- [2] N. Chernov, Stovan, Y., Romanova, T. "Mathematical model and efficient algorithms for object packing problem". *Comput. Geom.: Theory and Appl.*, 43(5), pp. 535–553 2010
- [3] Yu. Stoyan, T. Romanova "Mathematical Models of Placement Optimisation: Two- and Three-Dimensional Problems and Applications". In: *Fasano G, Pinter J. (eds.) "Modeling and Optimization in Space Engineering"*, 73, pp.363-388. Springer Optimization and its Applications, New York, 2012
- [4] A. Kovalenko, T. Romanova, P. Stetsyuk "Balance layout problem for 3D-objects: mathematical model and solution methods". *Cyb.and Syst. Anal.*, 51(4), pp. 556-565, 2015
- [5] Yu. Stoyan, T. Romanova, A. Pankratov, A. Kovalenko, P. Stetsyuk "Modeling and Optimization of Balance Layout Problems". In: *Fasano G, Pinter J. (eds.) "Space Engineering. Modeling and Optimization with Case Studies"*, 114, pp. 369-400. Springer Optimization and its Applications, New York, 2016
- [6] S. Butenko, P. Pardalos, V. Shylo "Optimization Methods and Applications". In: *Springer Optimization and Its Applications Series*, 130, 2017
- [7] C. Papadimitriou, K. Steiglitz "Combinatorial Optimization: Algorithms and Complexity". *Courier Corporation*, 1998
- [8] E. Reingold, J. Nievergelt, N. Deo "Combinatorial Algorithms: Theory and Practice". *Pearson Education, Canada*, 1977
- [9] V. Sachkov "Combinatorial Methods in Discrete Mathematics". *Cambridge University Press*, Edition 1, 1996.
- [10] Yu. Stoyan, I. Grebennik "Description and Generation of Combinatorial Sets Having Special Characteristics". *Inter. Jour. of Biomed. Soft Comp. and Hum. Scien., Spec. vol. Bilevel Programming, Optimization Methods, and Applications to Economics*, 18 (1), pp. 83-88, 2013
- [11] I. Grebennik "Description and generation of permutations containing cycles". *Cybern. Syst. Analysis*, 46(6), pp. 945-952, 2010
- [12] D. Knuth "The Art of Computer Programming, 4(2): Generating All Tuples and Permutations". Addison-Wesley, Boston, 2005
- [13] D. Kreher, D. Stinson "Combinatorial Algorithms: Generation, Enumeration and Search". *CRC Press*, 1999
- [14] F. Ruskey "Combinatorial Generation", *Dept. of Comput. Sci. Univ. of Victoria, Canada*, 1j-CSC 425/20, 2003
- [15] W. Lipski "Combinatorics for Programmers" [in Polish], *Polish Sci. Publ. (PWN)*, Warsaw, 1982
- [16] Yu. Stoyan, A. Pankratov, T. Romanova, A. Chugay "Optimized object packings using quasi-phi-functions". In: *Fasano G, Pinter J. (eds.) Optimized Packings and Their Applications*, 105, pp. 265-291. Springer Optimization and its Applications, New York, 2015
- [17] Yu. Stoyan, A. Pankratov, T. Romanova "Quasi phi-functions and optimal packing of ellipses". *J. of Glob. Optim.* 65 (2), pp. 283-307, 2016.

# Muticriteria Model of Balanced Layout Problem of 3D-Objects

Inna Urniaieva<sup>1</sup>, Igor Grebennik<sup>1</sup>, Tatiana Romanova<sup>2</sup>, Alexandr Pankratov<sup>2</sup>, Anna Kovalenko<sup>1</sup>

1. Department of Computer Science, Kharkiv National University of Radioelectronics, UKRAINE, Kharkiv, 14 Nauki ave., email: [igorgrebennik@gmail.com](mailto:igorgrebennik@gmail.com)

2. Department of Mathematical Modeling and Optimal Design, Institute for Mechanical Engineering Problems of the National Academy of Sciences of Ukraine, UKRAINE, Kharkiv, 2/10 Pozharskogo str., email: [sherom@kharkov.ua](mailto:sherom@kharkov.ua)

**Abstract:** The paper studies the optimal layout problem of 3D-objects. The problem takes into account placement constraints, as well as, behaviour characteristics of the mechanical system. We construct a mathematical model of the problem in the form of a multicriteria optimisation problem and call the problem Multicriteria Balanced Layout Problem (MBLP).

**Keywords:** Layout problem, Behaviour Constraints, Placement Constraints, Multicriteria Optimisation.

## I. INTRODUCTION

3D layout optimisation problems have a wide spectrum of practical applications. In particular, these problems arise in space engineering for rocketry design. Their distinctive feature consists of taking into account behaviour constraints of a satellite system. Behaviour constraints specify the requirements for system's mechanical properties such as equilibrium, inertia, and stability. Many publications analyze problems of the equipment layout in modules of spacecraft or satellites [1, 2]. These problems are NP-hard.

In the research we consider the balance layout problem in the following statement: arrange 3D-objects in a container taking into account special placement and behaviour constraints so that the objective function attains its extreme value [3].

We consider here an extension of the balanced layout of 3D-objects considered in [3,4]. The paper studies 3D optimisation balance layout problem taking into account minimal and maximal allowable distances. Classes of adjusted phi-functions and adjusted quasi-phi-functions are derived for analytical description of non-overlapping, containment and distance constraints. A circular cylinder, a paraboloid, or a truncated cone are taken as a container. We consider cylinders, spheres, tori, spherocylinders and straight convex prisms as the placement objects. An exact mathematical model of the problem in the form of NLP problem is provided.

The aim of this study is to develop a mathematical model of 3D layout optimisation problem taking into account behavior constraints in the form of multicriteria optimisation problem. We call the problem the Multicriteria Balanced Layout Problem (MBLP).

To describe placement constraints (non-overlapping of objects, containment of objects in a container with regard for the minimal and maximal allowable distances) analytically

we employ phi-function technique [5]. We also formalise behaviour constraints (equilibrium, moments of inertia, and stability constraints) based on [3].

The variety of forms of objective functions and combinations of placement and behaviour constraints generates various variants of the MBLP problem.

## II. PROBLEM FORMULATION

Let  $\Omega = \{(x, y, z) \in \mathbb{R}^3 : G(x, y, z) \geq 0\}$  be a container of given height  $H$ . We consider the following types of containers: 1)  $\Omega \equiv C$ ,  $C$  is a straight circular cylinder with a base of radius  $R$ ,  $G(x, y, z) = \min\{-x^2 - y^2 + R^2, -z + H, z\}$ ; 2)  $\Omega \equiv \Lambda$ ,  $\Lambda$  is a paraboloid of revolution with a base of radius  $R = \sqrt{H}$ ,  $G(x, y, z) = \min\{-z - x^2 - y^2 + H, z\}$ ; 3)  $\Omega \equiv E$ ,  $E$  is a straight circular blunted cone with lower and upper bases of radii  $R_1$  and  $R_2 < R_1$  respectively,  $G(x, y, z) = \min\{-z - H(\sqrt{x^2 + y^2} - R_1)/(R_1 - R_2), -z + H, z\}$ . Suppose that  $\Omega$  is divided by circular racks  $S_k$ ,  $k = 1, 2, \dots, m+1$ , into subcontainers  $\Omega^k$ ,  $k = 1, 2, \dots, m$ . We assume that  $S_1$  is a base of  $\Omega$ . Between racks  $S_k$  and  $S_{k+1}$  the distance  $t_k$  is given.

Family  $A = \{A_i, i \in I_n\}$ ,  $I_n = \{1, 2, \dots, n\}$ , involves the following shapes of objects: solid spheres  $S_i$  of radius  $r_i$ ; straight circular cylinders  $C_i$  of radius  $r_i$  and height  $2h_i$ ; tori  $T_i$  with metric characteristics  $(r_i, h_i)$ , where  $r_i$  is the distance from the center of generating circle to the axis of revolution,  $2h_i$  is the height of  $T_i$ ,  $h_i$  is the radius of the generating circle; spherocylinders  $S_{Ci}$  with metric characteristics  $(l_i, r_i, h_i)$ , where  $l_i$  is the height of ball segments,  $r_i$  is the radius and  $2h_i$  is the height of cylinder; straight regular prisms and cuboids  $K_i$  with metric characteristics  $(h_i, \tilde{v}_{il},)$ , where  $2h_i$  is the height of  $K_i$ ,  $\tilde{v}_{il} = (\tilde{x}_{il}, \tilde{y}_{il})$ ,  $l = 1, \dots, s_i$ , are vertices of the base of  $K_i$  (which is a convex polygon  $K_i$ ),  $s_i$  is the number of vertices of  $K_i$ .

**MBLP:** Pack 3D-objects  $A_i \in A$ ,  $i \in I_n = \{1, 2, \dots, n\}$ , inside container  $\Omega$ , so that the vector function attains its extreme value with regard for placement and behaviour constraints.

The placement constraints in the MBLP problem are generated by non-overlapping of objects  $A_i, A_j$ ,  $i > j \in I_n$ , which have to be placed inside container  $\Omega$ , and containment of object  $A_i$  in container  $\Omega$ ,  $i \in I_n$ . In addition, the minimal  $\rho_{ij}^-$  and maximal  $\rho_{ij}^+ \geq \rho_{ij}^-$  allowable distances between objects  $A_i, A_j$ ,  $i > j \in I_n$ , may be specified. Also, the minimal allowable distance  $\rho_i^-$  between object  $A_i \in A$ ,  $i \in I_n$ , and the lateral surface of container  $\Omega$  may be given. Without loss of generality we set  $\rho_{ij}^- = 0$  (or  $\rho_{ij}^+ = \varpi$ ) if a minimal (or a maximal) allowable distance between objects  $A_i$  and  $A_j$  is not given,  $i > j \in I_n$ . Here  $\varpi$  is a given sufficiently great number. In particular, the condition  $\rho_{ij}^+ = \rho_{ij}^-$  provides the arrangement of objects  $A_i$  and  $A_j$  on the exact distance. We also set  $\rho_i^- = 0$  if a minimal allowable distance between object  $A_i$  and the lateral surface of container  $\Omega$  is not given.

Placement constraints in the MBLP problem may be presented as the following:  $\rho_{ij}^- \leq \text{dist}(A_i, A_j) \leq \rho_{ij}^+$ ,  $i > j \in I_n$ , and  $\text{dist}(A_i, \Omega^*) \geq \rho_i^-$ ,  $i = 1, \dots, n$ , where  $\Omega^* = \mathbb{R}^3 \setminus \text{int } \Omega$ .

To describe the placement constraints analytically we employ the phi-function technique [5–8].

Let us consider the constraints of mechanical characteristics of system  $\Omega_A$ .

*The equilibrium constraints* are defined by the following system of inequalities:

$$\begin{aligned}\mu_{11}(u) &= \min\{-(x_s(u) - x_e) + \Delta x_e, (x_s(u) - x_e) + \Delta x_e\} \geq 0 \\ \mu_{12}(u) &= \min\{-(y_s(u) - y_e) + \Delta y_e, (y_s(u) - y_e) + \Delta y_e\} \geq 0 \\ \mu_{13}(u) &= \min\{-(z_s(u) - z_e) + \Delta z_e, (z_s(u) - z_e) + \Delta z_e\} \geq 0,\end{aligned}$$

where  $(x_e, y_e, z_e)$  is the expected position of  $O_s$ ,  $(\Delta x_e, \Delta y_e, \Delta z_e)$  are admissible deviations from the point  $(x_e, y_e, z_e)$ .

*The constraints of moments of inertia* are defined as the following:

$$\begin{aligned}\mu_{21}(u) &= -J_X(u) + \Delta J_X \geq 0, \\ \mu_{22}(u) &= -J_Y(u) + \Delta J_Y \geq 0, \\ \mu_{23}(u) &= -J_Z(u) + \Delta J_Z \geq 0,\end{aligned}$$

where  $J_X(u), J_Y(u), J_Z(u)$  are the moments of inertia of the system  $\Omega_A$  with respect to the axes of coordinate system  $O_sXYZ$ ,  $\Delta J_X, \Delta J_Y, \Delta J_Z$  are admissible values for  $J_X(u), J_Y(u), J_Z(u)$ , where

$$\begin{aligned}J_X(u) &= J_{x_0} + \sum_{i=1}^n (J_{x_i} \cos^2 \theta_i + J_{y_i} \sin^2 \theta_i) + \sum_{i=1}^n (y_i^2 + z_i^2)m_i - M(y_s^2 + z_s^2), \\ J_Y(u) &= J_{y_0} + \sum_{i=1}^n (J_{x_i} \sin^2 \theta_i + J_{y_i} \cos^2 \theta_i) + \sum_{i=1}^n (x_i^2 + z_i^2)m_i - M(x_s^2 + z_s^2) \\ J_Z(u) &= \sum_{i=0}^n J_{z_i} + \sum_{i=1}^n (y_i^2 + z_i^2)m_i - M(x_s^2 + y_s^2),\end{aligned}$$

$J_{x_0}, J_{y_0}, J_{z_0}$  are the moments of inertia of container  $\Omega$  with respect to the axes of the coordinate system  $Oxyz$ ,  $J_{x_i}, J_{y_i}, J_{z_i}$ ,  $i \in I_n$ , are the moments of inertia of object  $A_i$  with respect to the axes of coordinate system  $O_i x_i y_i z_i$ .

*The stability constraints* are defined by the following system of inequalities:

$$\mu_{31}(u) = \min\{-J_{XY}(u) + \Delta J_{XY}, J_{XY}(u) + \Delta J_{XY}\} \geq 0,$$

$$\mu_{32}(u) = \min\{-J_{YZ}(u) + \Delta J_{YZ}, J_{YZ}(u) + \Delta J_{YZ}\} \geq 0,$$

$$\mu_{33}(u) = \min\{-J_{XZ}(u) + \Delta J_{XZ}, J_{XZ}(u) + \Delta J_{XZ}\} \geq 0,$$

where  $J_{XY}(u), J_{YZ}(u), J_{XZ}(u)$  are the products of inertia of system  $\Omega_A$  with respect to the axes of the coordinate system  $O_sXYZ$ ,  $\Delta J_{XY}, \Delta J_{YZ}, \Delta J_{XZ}$  are admissible values for  $J_{XY}(u), J_{YZ}(u), J_{XZ}(u)$ , respectively,

$$\begin{aligned}J_{XY}(u) &= \frac{1}{2} \sum_{i=1}^n (J_{x_i} - J_{y_i}) \sin 2\theta_i + \sum_{i=1}^n x_i y_i m_i - M x_s y_s, \\ J_{YZ}(u) &= \sum_{i=1}^n y_i z_i m_i - M y_s z_s, \\ J_{XZ}(u) &= \sum_{i=1}^n x_i z_i m_i - M x_s z_s.\end{aligned}$$

*Behaviour constraints* of the BLP problem we define as the system of inequalities

$$\mu_1(u) \geq 0, \mu_2(u) \geq 0, \mu_3(u) \geq 0,$$

where

$$\mu_1(u) = \min\{\mu_{11}(u), \mu_{12}(u), \mu_{13}(u)\}, \quad (1)$$

$$\mu_2(u) = \min\{\mu_{21}(u), \mu_{22}(u), \mu_{23}(u)\}, \quad (2)$$

$$\mu_3(u) = \min\{\mu_{31}(u), \mu_{32}(u), \mu_{33}(u)\}. \quad (3)$$

Here  $O_s = (x_s, y_s, z_s)$  is the center of mass of system  $\Omega_A$ ,

$$x_s(u) = \frac{1}{M} \sum_{i=1}^n m_i x_i, \quad y_s(u) = \frac{1}{M} \sum_{i=1}^n m_i y_i,$$

$$z_s(u) = \frac{1}{M} \sum_{i=1}^n m_i z_i, \quad M = \sum_{i=1}^n m_i \text{ is the mass of system } \Omega_A.$$

### III. MATHEMATICAL MODEL

A mathematical model of the MBLP problem can be presented in the form

$$\text{extr}F(p, u) \text{ s.t. } (u, p) \in W \quad (4)$$

$$W = \{(u, p) \in \mathbb{R}^\xi : \Upsilon(u, p) \geq 0, \mu(u, p) \geq 0, \zeta \geq 0\}, \quad (5)$$

where  $F(p, u) = (F_1(p, u), F_2(p, u), \dots, F_k(p, u))$ ,  
 $\Upsilon(u, p)$  describes placement constraints,  
 $\Upsilon(u, p) = \min\{\Upsilon_1(u), \Upsilon_2(u, p)\}$ ,  
 $\Upsilon_1(u)$  is responsible for non-overlapping constraints,  
 $\Upsilon_2(u, p)$  is responsible for containment constraints,  
 $\mu(u) = \min\{\mu_s(u), s \in U_t\}$  is responsible for behavior constraints,  $U_t \in P(U)$ ,  $P(U)$  is the power set of  $U = \{1, 2, 3\}$ , functions  $\mu_1(u), \mu_2(u), \mu_3(u)$  are given by (1)-(3),  $\zeta \geq 0$  is the system of additional constraints of metric characteristics of container  $\Omega$  and placement parameters of objects. If  $s = \emptyset$ , i.e. behavior constraints are not involved in (5), then our objective function  $F(u)$  meets mechanical characteristics of system  $\Omega_A$ .

Depending on the different combinations of objective functions  $F_1(p, u), F_2(p, u), \dots, F_k(p, u)$  different variants of mathematical model (4)-(5) can be generated. The most frequently occurring objective functions found in related publications are the following: 1) size of container  $\Omega$ ; 2) deviation of the center of mass of system  $\Omega_A$  from a given point; 3) moments of inertia of system  $\Omega_A$  [3,9-13].

Let us consider some of realisations of model (4) - (5):

- $F(p, u) = p$  s.t.  $(p, u) \in W \subset \mathbb{R}^\xi$ ,

$$W = \{(p, u) \in \mathbb{R}^\xi : \Upsilon_1(u) \geq 0, \Upsilon_2(p, u) \geq 0, \mu(p, u) \geq 0, \zeta \geq 0\}$$

- $F(u) = d$ ,  $(p, u) \in W \subset \mathbb{R}^\xi$ ,

$$d = (x_s(u) - x_e)^2 + (y_s(u) - y_e)^2 + (z_s(u) - z_e)^2,$$

$$W = \{(p, u) \in \mathbb{R}^\xi : \Upsilon_1(u) \geq 0, \Upsilon_2(p, u) \geq 0, \mu_2(p, u) \geq 0, \mu_3(p, u) \geq 0, \zeta \geq 0\};$$

- $F(p, u) = (F_1(p, u) = p, F_2(p, u) = d)$ ,  $(p, u) \in W \subset \mathbb{R}^\xi$ ,

$$W = \{(p, u) \in \mathbb{R}^\xi : \Upsilon_1(u) \geq 0, \Upsilon_2(p, u) \geq 0, \mu_2(p, u) \geq 0, \mu_3(p, u) \geq 0, \zeta \geq 0\};$$

- $F(p, u) = (F_1(p, u) = J_X(p, u), F_2(p, u) = J_Y(p, u),$

$$F_3(p, u) = J_Z(p, u))$$
  

$$(p, u) \in W \subset \mathbb{R}^\xi,$$

$$W = \{(p, u) \in \mathbb{R}^\xi : \Upsilon_1(u) \geq 0, \Upsilon_2(p, u) \geq 0, \mu_1(p, u) \geq 0, \mu_3(p, u) \geq 0, \zeta \geq 0\}.$$

### III. CONCLUSIONS

In this paper we formulate the optimisation layout problem of 3D-objects into a container taking into account placement (non-overlapping, containment, distance) and behaviour (equilibrium, inertia and stability) constraints. We call the problem as Multicriteria Balance Layout Problem (MBLP). In order to describe placement constraints analytically we employ phi-function technique. A mathematical model of the problem in the form of multicriteria optimisation problem is proposed.

We also consider some variants of the MBLP problem

depending on the forms of the objective functions, shapes of objects and containers, combinations of placement and behavior constraints.

### IV. COMPUTATIONAL RESULTS

**Instance 1.** Let  $\Omega \equiv \mathbf{E}$ ,  $m = 2$ ,  $H = 0.6$ ,  $R_1 = 0.5$ ,  $R_3 = 0.3$ ,  $A = \{\mathcal{S}_1, \mathcal{S}_2, \mathcal{C}_3, \mathcal{T}_4, \mathcal{T}_5, \mathcal{S}_{C7}, \mathcal{S}_{C8}, \mathcal{K}_9, \mathcal{K}_{10}\}$ ,  $A_-^1 = \{\mathcal{S}_1, \mathcal{C}_3, \mathcal{T}_5, \mathcal{S}_{C7}, \mathcal{K}_9\}$ ,  $A_+^2 = \{\mathcal{S}_2, \mathcal{C}_4, \mathcal{T}_6, \mathcal{S}_{C8}, \mathcal{K}_{10}\}$ ,  $\rho_{ij}^- = 0.03$ ,  $i < j \in I_{10}$ ,  $\rho_{39}^+ = 0.1$ ,  $\rho_{26}^+ = 0.08$ ,  $(x_e, y_e, z_e) = (0, 0, 0.275)$ ,  $t_1 = 0.3$ ,  $n = 10$ ,  $\{z_i, i = 1, \dots, 10\} = \{0.19, 0.4, 0.19, 0.41, 0.24, 0.35, 0.19, 0.39, 0.18, 0.42\}$ ,  $\{m_i, i = 1, \dots, 10\} = \{27.8764, 20.944, 34.5575, 16.9332, 28.4245, 22.2066, 17.2159, 19.2265, 38.4, 19.9532\}$ ,  $r_1 = 0.11$ ,  $r_2 = 0.1$ ,  $r_3 = 0.1$ ,  $h_3 = 0.11$ ,  $r_4 = 0.07$ ,  $h_4 = 0.11$ ,  $r_5 = 0.08$ ,  $h_5 = 0.06$ ,  $r_6 = 0.09$ ,  $h_6 = 0.05$ ,  $r_7 = 0.08$ ,  $h_7 = 0.05$ ,  $l_7 = 0.06$ ,  $h_8 = 0.06$ ,  $l_8 = 0.03$ ,  $s_9 = 4$ ,  $h_9 = 0.12$ ,  $\tilde{v}_{91} = (0.08, 0.1)$ ,  $\tilde{v}_{92} = (0.08, -0.1)$ ,  $\tilde{v}_{93} = (-0.08, -0.1)$ ,  $\tilde{v}_{94} = (-0.08, 0.1)$ ,  $s_{10} = 6$ ,  $h_{10} = 0.12$ ,  $\tilde{v}_{(10)1} = (0.04, 0.07)$ ,  $\tilde{v}_{(10)2} = (0.08, 0)$ ,  $\tilde{v}_{(10)3} = (0.04, -0.07)$ ,  $\tilde{v}_{(10)4} = (-0.04, -0.07)$ ,  $\tilde{v}_{(10)5} = (-0.08, 0)$ ,  $\tilde{v}_{(10)6} = (-0.04, 0.07)$ .

The local-optimal solution found by NLP-solver in CAS Math 9 (Fig. 1) is  $F(u^*, u'^*) = 1.12726 \times 10^{-6}$ .

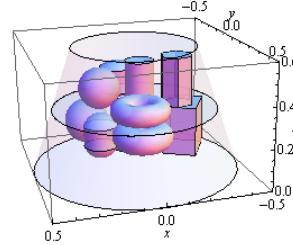


Fig. 1 The local optimal layout of 3D-objects in Instance 1

**Instance 2.** Let  $\Omega \equiv \mathbf{C}$ ,  $m = 3$ ,  $H = 1$ ,  $R = 0.45$ ,  $t_2 = 0.35$ ,  $n = 20$ ,  $A = \{\mathcal{S}_i, i = 1, \dots, 4, \mathcal{C}_i, i = 5, \dots, 8, \mathcal{T}_i, i = 9, \dots, 12, \mathcal{S}_{C_i}, i = 13, \dots, 16, \mathcal{K}_i, i = 17, \dots, 20\}$ ,  $A_-^1 = \{\mathcal{S}_1, \mathcal{C}_5, \mathcal{C}_6, \mathcal{T}_9, \mathcal{S}_{C14}, \mathcal{P}_{17}\}$ ,  $A_+^2 = \{\mathcal{S}_2, \mathcal{S}_3, \mathcal{C}_7, \mathcal{T}_{10}, \mathcal{S}_{C15}, \mathcal{P}_{18}, \mathcal{K}_{20}\}$ ,  $A_+^3 = \{\mathcal{S}_4, \mathcal{C}_8, \mathcal{T}_{11}, \mathcal{T}_{12}, \mathcal{S}_{C16}, \mathcal{P}_{19}\}$ ,  $U_t = \emptyset$ ,  $\rho_{ij}^- = 0.02$ ,  $i < j = 1, \dots, 20$ ,  $(x_e, y_e, z_e) = (0, 0, 0.5)$ ,  $\{z_i, i = 1, \dots, 20\} = \{0.1, 0.44, 0.46, 0.81, 0.11, 0.12, 0.46, 0.78, 0.06, 0.425, 0.76, 0.77, 0.11, 0.13, 0.46, 0.81, 0.12, 0.47, 0.82, 0.46\}$ ,  $\{m_i, i = 1, \dots, 20\} = \{20.944, 15.2681, 27.8764, 34.5575, 63.7115, 41.8146, 30.4106, 28.4245, 49.9649, 24.8714, 38.6888, 26.2637, 20.7764, 17.2159, 16.8756, 52.8, 52.8, 52.8, 23.1489\}$ ,  $r_1 = 0.1$ ,  $r_2 = 0.09$ ,  $r_3 = 0.11$ ,

$$\begin{aligned}
& r_4 = 0.11, \quad r_5 = 0.1, \quad h_5 = 0.11, \quad h_6 = 0.12, \quad r_7 = 0.11, \\
& r_8 = 0.11, \quad h_8 = 0.08, \quad r_9 = 0.08, \quad h_9 = 0.07, \quad r_{10} = 0.09, \\
& h_{10} = 0.075, \quad r_{11} = 0.07, \quad h_{11} = 0.06, \quad r_{12} = 0.08, \\
& h_{12} = 0.07, \quad r_{13} = 0.1, \quad h_{13} = 0.05, \quad l_{13} = 0.07, \quad r_{14} = 0.05, \\
& h_{14} = 0.05, \quad l_{14} = 0.08, \quad r_{15} = 0.08, \quad h_{15} = 0.05, \quad l_{15} = 0.06, \\
& r_{16} = 0.08, \quad h_{16} = 0.04, \quad l_{16} = 0.07, \quad s_i = 4, \\
& \tilde{v}_{i1} = (-0.11, -0.1), \quad \tilde{v}_{i2} = (0.11, -0.1), \quad \tilde{v}_{i3} = (0.11, 0.1), \\
& \tilde{v}_{i4} = (-0.11, 0.1), \quad h_i = 0.12, \quad i = 17, 18, 19, \quad s_{20} = 6, \\
& \tilde{v}_{(20)1} = (0.045, 0.078), \quad \tilde{v}_{(20)2} = (0.09, 0), \quad \tilde{v}_{(20)3} = (0.045, -0.078), \\
& \tilde{v}_{(20)4} = (-0.045, -0.078), \quad \tilde{v}_{(20)5} = (-0.09, 0), \\
& \tilde{v}_{(20)6} = (-0.045, 0.078), \quad h_{20} = 0.11.
\end{aligned}$$

The local optimal solution found by NLP-solver in CAS Math 9 is  $F(u^*, u'^*) = 0.001911$  (Fig. 2).

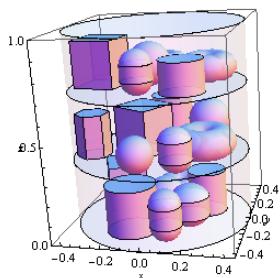


Fig. 2 The local optimal layout of 3D-objects in Instance 2

81, 54, 29, 94, 92, 41, 57, 77, 40, 67, 31, 47, 39, 61, 73, 83, 11, 20, 86, 72, 81, 54, 29, 94, 92, 41, 57, 77, 40, 67, 31, 47, 39, 61, 73, 83, 11, 20, 86, 72, 81, 54, 29, 94, 92, 41, 57, 77, 40, 67, 31, 47, 39, 61, 73, 83, 11, 20, 86, 72, 81, 54, 29, 94, 92, 41, 57, 77, 40, 67, 31, 47, 39, 61, 73, 83, 11, 20}.

The local optimal solution found by IPOPT is  $F(u^*, u') = 0.000000$  (Fig. 3).

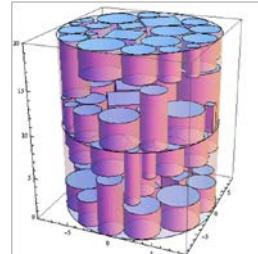


Fig. 3 The local optimal layout of 3D-objects in Instance 3

## REFERENCES

- [1] G. Fasano, J. Pinte'r (Eds.) "Modeling and Optimisation in Space Engineering" Series: *Springer Opt. and Its Appl.* 73, XII.404, 2013
  - [2] G. Fasano, J. Pinte'r (Eds.) "Optimized Packings and Their Applications". *Springer Opt. and its Appl.* 105, 326 2015.
  - [3] C. Che, Y. Wang, H. Teng "Test problems for quasi-satellite packing: Cylinders packing with behaviour constraints and all the optimal solutions known". *Opt. Online* [http://www.optimisation-online.org/DB\\_HTML/2008/09/2093.html](http://www.optimisation-online.org/DB_HTML/2008/09/2093.html), 2008.
  - [4] Yu. Stoyan, T. Romanova, A. Pankratov, A. Kovalenko, P. Stetsyuk "Modeling and Optimization of Balance Layout Problems". In: *Fasano G., Pinter J. (eds.) "Space Engineering. Modeling and Optimization with Case Studies"*, 114, pp. 369-400. Springer Optimization and its Applications, New York, 2016
  - [5] Yu. Stoyan, T. Romanova "Mathematical Models of Placement Optimisation: Two- and Three-Dimensional Problems and Applications". In: *Fasano G., Pinter J. (eds.) "Modeling and Optimization in Space Engineering"*, 73, pp.363-388. Springer Optimization and its Applications, New York, 2012
  - [6] Yu. Stoyan, A. Pankratov, T. Romanova "Quasi phi-functions and optimal packing of ellipses". *J. of Glob. Optim.* 65 (2), pp. 283-307, 2016
  - [7] Yu. Stoyan, T. Romanova, A. Pankratov, A. Kovalenko, P. Stetsyuk "Modeling and Optimization of Balance Layout Problems". In: *Fasano G., Pinter J. (eds.) "Space Engineering. Modeling and Optimization with Case Studies"*, 114, pp. 369-400. Springer Optimization and its Applications, New York, 2016
  - [8] Yu. Stoyan, T. Romanova, A. Pankratov, A. Chugay "Optimized object packings using quasi-phi-functions". *Optimized Packings with Applications*, G. Fasano, J. D. Pinte'r (Eds.). Springer, New York, Vol. 105, 2015, pp. 265-293.
  - [9] Z. Sun, , H. Teng, "Optimal layout design of a satellite module". *Eng. Opt.* 35(5), pp. 513-530, 2003.
  - [10] L. Jingfa, L. Gang "Basin filling algorithm for the circular packing problem with equilibrium behavioural constraints", *SCIENCE CHINA Inf. Sci.*, 53(5), pp. 885-895, 2010.
  - [11] W.A. Oliveira, A.C. Moretti, L.L. Salles-Neto "A heuristic for the nonidentical circle packing problem", *Anais do CNMAC*, 3, pp. 626-632, 2010.
  - [12] Y.-C. Xu, , R.-B. Xiao, M. Amos "A novel algorithm for the layout optimisation problem", *Proc. 2007 IEEE Congr. Evolut. Comput. (CEC07)*, IEEE Press, pp. 3938-3942, 2007.
  - [13] A. Kovalenko, T. Romanova, P. Stetsyuk "Balance layout problem for 3D-objects: mathematical model and solution methods". *Cybern. Syst. Anal.* 51(4), pp. 556-565, 2015

# Statistical Analysis of Measuring Errors the Pollution of the Atmospheric Bottom Layer by Exhaust Gas

Iryna Darmorost<sup>1</sup>, Petro Stakhiv<sup>2</sup>, Mykola Shynkaryk<sup>3</sup>, Vasylyna Melnyk<sup>4</sup>

1,3,4. Department of Computer Science, Ternopil National Economic University, UKRAINE, Ternopil,  
8 Chekhova str., email: [imadiudia@gmail.com](mailto:imadiudia@gmail.com)<sup>1</sup>

2. Lodz University of Technology, POLAND, Lodz,  
116 Stefana Źeromskiego str., email: [spp@polynet.lviv.ua](mailto:spp@polynet.lviv.ua)

**Abstract:** The statistical analysis of the errors of measurements pollution in the bottom-layer of the atmosphere with exhaust gases of vehicles has been carried out. The method of performing statistical analysis is proposed. On this basis, the most optimal time averaging interval for the instantaneous concentrations of harmful substances is established.

**Keywords:** traffic intensity, harmful emissions, nitrogen dioxide, correlation coefficient, random error, averaging interval.

## I. INTRODUCTION

The progress of human society is inseparable from the history of transport development. With the expansion of states, the construction of cities, with the resettlement of people in increasingly large territories, the growth of trade rates of development of transport has steadily increased. The rapid pace of growth in the level motorization of the population leads to inevitable negative changes in the environment. Pollution of the atmosphere by harmful emissions of vehicles causes irreparable harm to the health of the population. It is known that during the day, people consume about 15-25 kg of air, 2.5-5 kg of water, 2.5 kg of food. When inhaled, the chemical elements are absorbed by the body most intensively. Thus, lead, which comes with air, is absorbed by blood by 60%, if it comes with water, it is absorbed by 10%, with food - by 5%. Therefore, when polluting the environment, atmospheric air is the main supplier of toxic substances in the human body [1].

The exhaust gases of cars contain more than 200 compounds and components, many of which are very toxic.

The environment contains carbon monoxide (CO), nitrogen (NO), sulfur dioxide (SO<sub>2</sub>), aldehydes, soot (C), lead (Pb), and others [2]. One of the most harmful substances, the concentration of which exceeds any allowable norms is nitrogen dioxide, the chemical formula of which – NO<sub>2</sub>.

There is a clear linear dependence between the traffic intensity and the concentration of harmful substances in the atmosphere. However, measuring the value of NO<sub>2</sub>, with the use of modern means of measurement, occurs almost instantaneously, and in this connection, the problem of averaging the obtained indicators. In the paper [3], experimental studies were performed and the correlation coefficients were calculated for different time averaging intervals: for 5, 10, 20 and 30 minutes. The correlation coefficient is the most optimal at the averaging interval in 20

minutes. In this work it is proposed to investigate the statistical characteristics of random error, in order to confirm the effectiveness of the selected averaging interval.

## II. STATEMENT OF THE PROBLEM

The traffic intensity is one of the most important factors contributing to the pollution of the bottom-layer of the atmosphere by harmful emissions, but the concentration of air in a chemical compound such as nitrogen dioxide varies depending on the characteristics of the medium in which the measurements are carried out.

Based on preliminary studies, it was found that the traffic intensity and the known concentrations of harmful substances in the atmosphere in certain time bands correlate. The research was based on the assumption that the random error is normally distributed with zero mathematical expectation, and as a result of averaging, is compensated for this error.

And on this basis the idea was based on the choice of averaging interval.

It has been established that in order to compensate for the casual component related to the influence of other factors on pollution of the city's territory, such as ventilation, it is advisable to choose the range of averaging instantaneous values of the concentration of harmful emissions and the traffic intensity of the transport units close to the interval of 20 minutes. Such an interval is used in standard measuring techniques Sanitary and Epidemiological Services.

In order to confirm or refute the justified averaging interval, it is proposed to investigate the statistical characteristics of the random error.

As is known, the correlation coefficient between the two variables is given by Eq.1:

$$r_{xy} = \frac{\sum_{i=1}^m (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^m (x_i - \bar{x})^2 \sum_{i=1}^m (y_i - \bar{y})^2}}, \quad (1)$$

where  $\bar{x}, \bar{y}$  – average sample from each sample  $\bar{x} = x_1 \dots x_m; \bar{y} = y_1 \dots y_m$  accordingly,  $r_{xy} \in [-1, 1]$

It is worth mentioning that the values  $\bar{x} = x_1 \dots x_m; \bar{y} = y_1 \dots y_m$ , are calculated for different time averaging intervals of instantaneous values of traffic intensity and concentration of harmful substances, respectively.

Thus, for different time averaging intervals of instantaneous values, we calculate the correlation coefficient between the traffic intensity and the concentration of nitrogen dioxide and carry out a statistical analysis of the errors measurements of contamination the bottom-layer of the atmosphere by exhaust gases of vehicles.

## II. METHOD OF RESEARCHING

Statistical analysis of the measurement errors the bottom-layer of atmosphere on the example of the established correlation between the traffic intensity and the concentration of nitrogen dioxide ( $\text{NO}_2$ ) was conducted.

To obtain experimental data, measurements of the concentration of nitrogen dioxide in the air at the crossroads of streets in city Ternopil Chekhova - Za Rudkoiu was conducted. Average air temperature was  $0^\circ\text{C}$ , humidity - 70%. Obtaining a sample of data was done using gas analyzer SPEC Sensors, DSG - NO2- 968-037, accuracy -  $\pm 15\%$ , at a range of operating temperatures from  $-20^\circ\text{C}$  to  $+40^\circ\text{C}$ . The traffic intensity was evaluated using a set webcam every minute. Concentration value  $\text{NO}_2$  received every second.

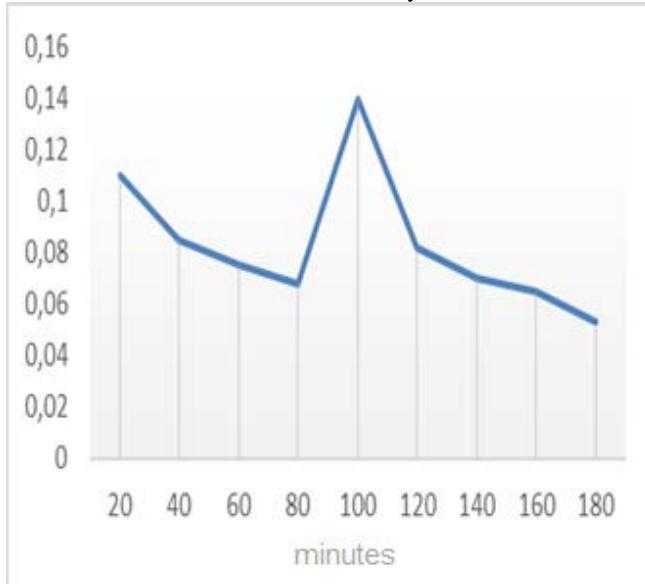


Fig.1. Interval of averaging concentration of  $\text{NO}_2$  in 20 minutes.

Using eq. (1), the correlation coefficients were calculated on time averaging intervals of 5, 10, 20, and 30 minutes. Figure 1 shows a graph of the averaging interval of nitrogen dioxide concentration in 20 minutes [3].

Figure 2 shows the generalized results of the correlation coefficients at different time intervals.

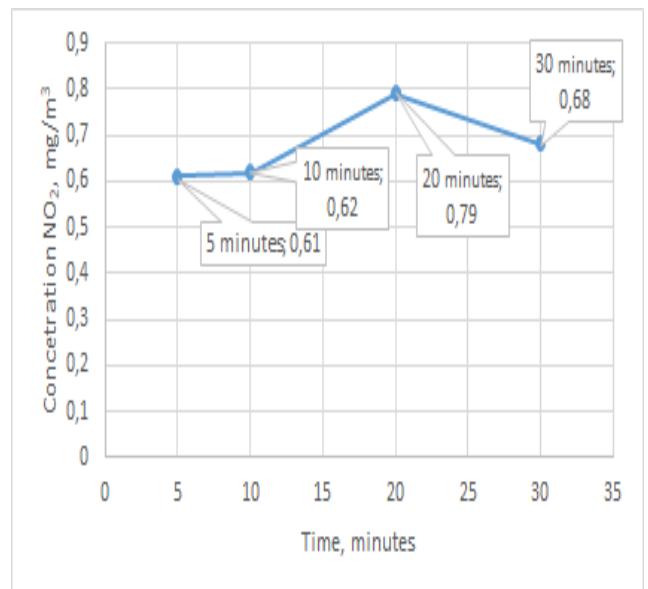


Fig.2. Coefficients of correlation for different time averaging intervals.

As you can see from Figure 2, we get a curve that has an extremum at the point with averaging of 20 minutes. For statistical analysis, the averaging intervals of 1, 5, 10, 20 and 30 minutes were selected.

The method of conducting the analysis includes the following steps:

1. Averaging of the concentrations of nitrogen dioxide in the bottom-layer of atmosphere at intervals of 1, 5, 10, 20, 30 minutes, in a time interval of 1 hour. For each interval averaging according to the value of the interval. For each of the intervals, 60, 12, 6, 3, and 2 averaging values were obtained, respectively.
2. On each averaging interval, net error and average value were selected. The calculation of the values of a random error occurs according to the following equations:

$$e_i = a_{mid} - z_m, \quad (2)$$

where  $e_i$  - random errors at different time intervals,  $a_{mid}$  - averaging the concentration of nitrogen dioxide in different time ranges,  $z_m$  - value measurement of nitrogen dioxide per second,  $m=1..n$ .

3. At all averaging intervals after calculations using eq. (2), the sum of random errors was obtained by the equations:

$$S_i = \sum_{i=1}^N e_i, \quad (3)$$

where  $S_i$  - amount of error values at different time averaging intervals.

## III. ANALYSIS OF RECEIVED RESULTS

The results of the research are shown in Figures 3-7. Statistical analysis of measurement errors, according to the

research method, was conducted for different time intervals, in particular: 1, 5, 10, 20 and 30 minutes.

Figure 3 shows the value of random errors in the averaging interval of 1 minute.

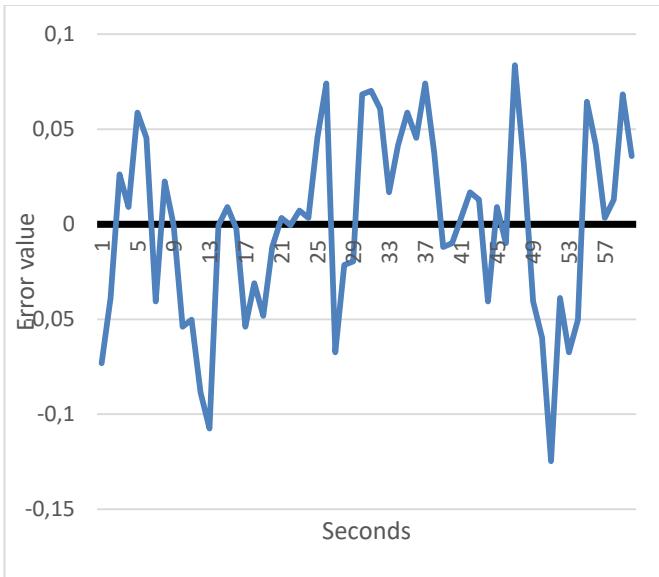


Fig. 3. The value of random errors in the time range of 1 minute in the averaging interval in 1 minute.

Figure 4 shows the value of random errors in the averaging interval in 5 minutes.

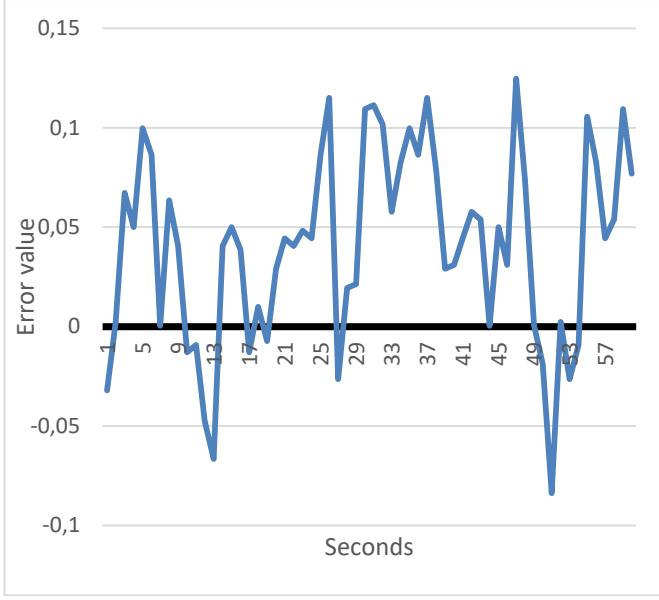


Fig. 4. The value of random errors in the time range of 1 minute in the averaging interval of 5 minutes.

Figure 5 shows the values of random errors in the averaging interval of 10 minutes.

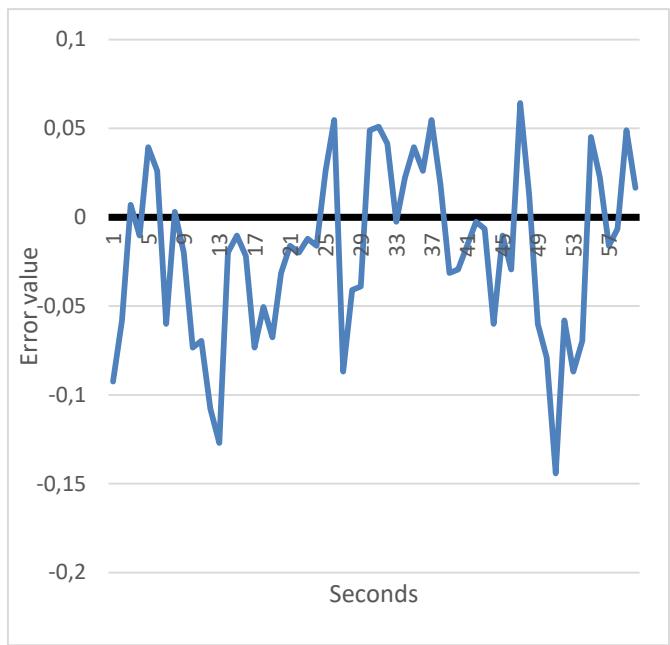


Fig. 5. The value of random errors in the time range of 1 minute in the averaging interval in 10 minutes.

Figure 6 shows the value of random errors in the averaging interval in 20 minutes.

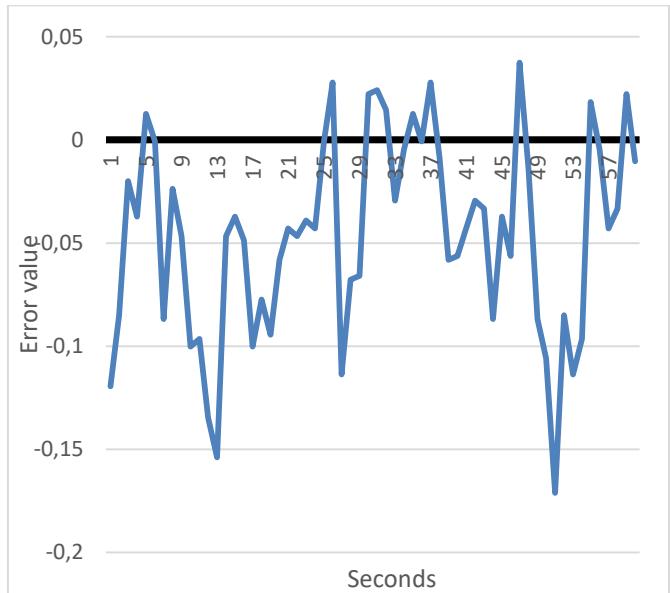


Fig. 6. The value of random errors in the time range of 1 minute in the averaging interval in 20 minutes.

Figure 7 shows the value of random errors in the averaging interval in 30 minutes.

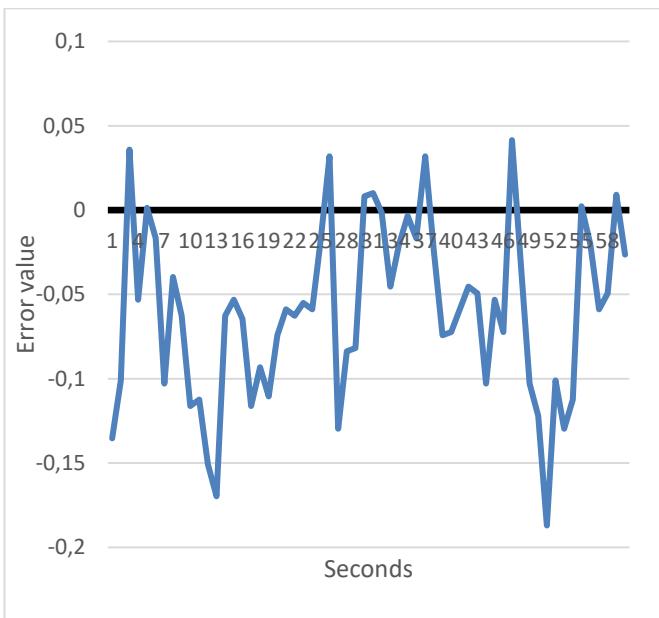


Fig. 7. The value of random errors in the time range of 1 minute in the averaging interval in 30 minutes.

On the basis of the proposed error, using the eq. (3) the general errors are calculated.

On the interval in 1 minute  $S_1 = -7,015$ , it means that the interval error is not normally distributed, but includes trend values, on 5 minutes  $S_5 = 13,148$ , which is also not the optimum value, in 10 minutes -  $S_{10} = 2,94$ , on 20 minutes -  $S_{20} = 2,808$ , on 30 minutes -  $S_5 = 3,0056$ .

Apparently, as a result of the calculations performed by the proposed methodology, it has been found that in subsequent studies, as averaging interval of , it is necessary to use the averaging interval in the range of 10 to 20 minutes.

The vast majority of sources [4] use an interval of averaging of 1 hour. The averaging graph for 1 hour is depicted in Figure 8. For the visualization, a sample of data was obtained from the measurement of nitrogen dioxide concentration values at Street Jana Pawla II in Lodz, Poland. On the ordinates axis, the averaged values are deferred value of nitrogen dioxide ( $\text{NO}_2$ ), along the abscissa axis - hours of day.

The conducted studies have shown that such averaging gives distortion of results due to the use of an incorrect and ineffective averaging interval of the measured values of nitrogen dioxide, which can not be neglected by the action of random factors such as ventilation of the environment, that is, gusts of wind, vertical and horizontal streams, turbulence, and so on.

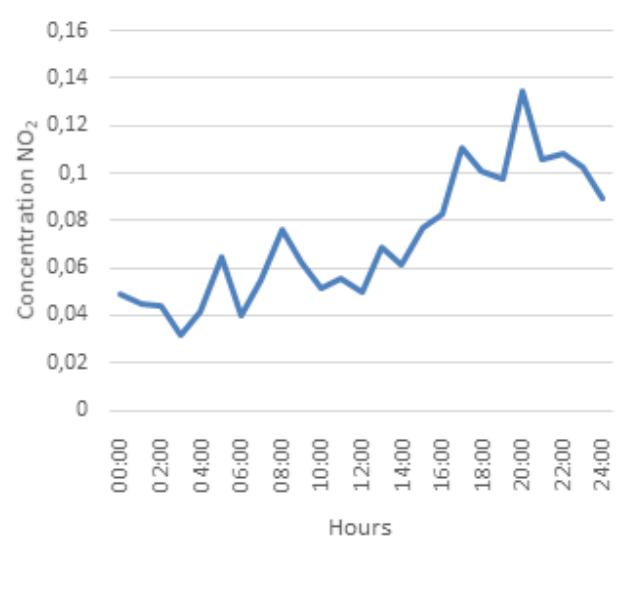


Fig. 8. Measured values of nitrogen dioxide in the averaging interval in 1 hour.

#### IV. ACKNOWLEDGMENT

This research was supported by National Grant of Ministry of Education and Science of Ukraine "Mathematical tools and software for control air pollution from vehicles" (0116U005507).

#### V. CONCLUSION

A statistical analysis of random errors was carried out at different intervals of averaging - at 1, 5, 10, 20 and 30 minutes. Thus, the averaging interval of instantaneous values of the concentration of nitrogen dioxide in bottom-layer of atmosphere was formed, which will be used in further studies to construct models of the dependence the concentration of harmful emissions between the traffic intensity . The value of the averaging interval will be in the range of 10 to 20 minutes.

#### REFERENCES

- [1] V. Stukanov, "Effect of motor transport on the state of the environment of the large industrial cities", *Vestnik VSU, Series: Chemistry, Biology, Pharmacy*, no.1, pp.168-175, 2012 (In Ukrainian).
- [2] V. Garin, "Industrial ecology", *Moscow-Marshrut*, 328 p., 2005 (In Russian).
- [3] M.Dyvak, I.Darmorost, R.Shevchuk, V.Manzhula and N.Kasatkina., "Correlation analysis traffic intensity of the motor vehicles and the air pollution by their harmful emissions", *Proceedings of XIVth International Conference on Modern Problem of Radio Engineering, Telecommunications and Computer Science (TCSET)*, Lviv-Slavskie, 2018, pp.855-858.
- [4] Wojewódzki Inspektorat Ochrony Środowiska w Łodzi [Online]. Available: <http://www.wios.lodz.pl/>. Accessed on: April, 22, 2018.

# A Mathematical Model of Microsurface Normal Distribution for Specular Bidirectional Reflectance Distribution Function

Ivan Dychka, Yevgeniya Sulema, Constantine Rudenko

Faculty of Applied Mathematics, Igor Sikorsky Kyiv Polytechnic Institute, Ukraine, Kyiv, 37 Peremohy ave.,  
email: {dychka, sulema}@pzks.fpm.kpi.ua, const.int@protonmail.com

**Abstract:** This paper presents a new mathematical model of microsurface normal distribution. The proposed model retains a long highlight tail even with roughness coefficient extremely close to zero which matches the behavior of measured material samples. The proposed model obeys the energy conservation law and can be used in Physically-Based Rendering (PBR) shaders to create realistic and visually pleasing specular highlights. The results of using the proposed model are demonstrated and discussed. The highlight shape created by this model is compared against several existing models as well as empirical data.

**Keywords:** Physically-Based Rendering, Specular Reflection, Normal Distribution, Realistic Rendering.

## I. INTRODUCTION

Nowadays there are several models of microsurface normal distribution available for specular Bidirectional Reflectance Distribution Function (BRDF) ranging from minimalist ones like Blinn-Phong to complex models like Cook-Torrance [1], Trowbridge-Reitz [2], Backmann [3], and other models. However, such models suffer from the same problem – with low roughness values the tail of the highlight is not nearly long enough even with the most advanced models.

The research presented in this paper is aimed at creating a new model of normal distribution for specular BRDF that would retain realistically long highlight tail with low roughness values.

## II. RELATED WORK

There is a relatively large number of research papers devoted to solving similar tasks.

Thus, Dupuy et al. proposed in [4] a reflectance filtering technique for displacement mapped surfaces called Linear Efficient Antialiased Displacement and Reflectance mapping. The authors assert that their method is compatible with animation and deformation, making it general and flexible.

Gartley et al. presented in [5] the framework for a new p-BRDF prediction tool leveraging the radiometric ray-tracing framework of the Digital Imaging and Remote Sensing Image Generation model. The authors state that predictions from the tool have been verified for clean, randomly rough surfaces against a generalized, analytical p-BRDF model.

In [6], Heitz and d'Eon proposed an importance sampling scheme for microfacet-based BSDFs. The authors claim that the performance of this scheme is suitable for both offline and GPU rendering.

Heitz in [7] presented the masking-shadowing functions (geometric attenuation factors) in microfacet-based BRDFs and provided explanations on their applications.

Hanika et al. in [8] proposed an advancement of microfacet theory based on the Smith model in order to include microsurface multiple scattering at rough material interfaces for reflectance and transmission. The authors compared the predictions made by their model with results obtained by simulating multiple scattering on explicit microsurfaces generated with a noise primitive.

Walteret et al. in [9] provided a review of the microfacet theory and demonstrated how it can be extended to simulate transmission through rough surfaces such as etched glass.

In [10], Dupuy et al. introduce the Symmetric GGX distribution to represent spatially-varying properties of anisotropic microflake participating media.

Despite of availability of these and other promising results a need of new models development still exists.

## III. MODEL DESCRIPTION

The proposed model is based on an assumption that the surface contains cavities of different depths with deeper cavities being less probable. In order to calculate the brightness of a given surface the following steps are performed:

- Calculation of the depth of a cavity that would reflect given light direction vector to the direction of given view vector;
- Calculation of the probability of required cavity.

The reflection of light vector which hits a cavity is demonstrated in Fig. 1.

The height of a cavity which reflects a given light vector into view vector is calculated using following the formula:

$$h = \sqrt{\frac{1}{b^2} - 1} \quad (1)$$

where  $h$  is the height of the cavity and  $b$  is the Blinn term.

The plot of cavity height function in relation to Blinn term is presented in Fig. 2.

After calculation of the cavity height, the probability distribution must be applied. Because Gaussian Distribution

Function does not provide the desired result, the proposed model uses a custom distribution function:

$$f(h, r) = \frac{r}{r+h} \quad (2)$$

where  $f$  is the distribution coefficient,  $h$  is the height of the cavity, and  $r$  is the surface roughness coefficient. This function is not normalized (does not integrate to 1) and, thus, should be divided by its integral to be used as a probability distribution.

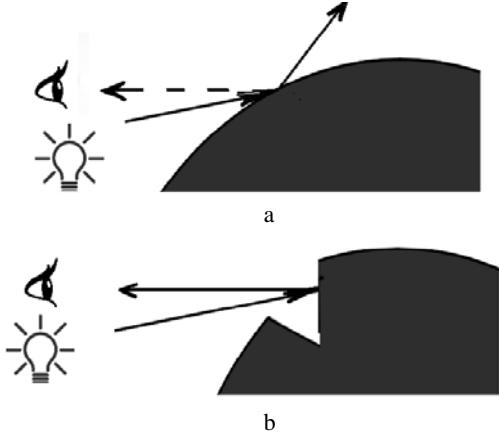


Fig. 1. Reflection cases: a – outside a cavity, b – inside a cavity

However, the integral of this function over  $h \in (0; \infty)$  does not converge. But instead it is possible to integrate the total amount of reflected light over  $\phi \in (0; \pi)$  where  $\phi$  is the angle between the cavity normal and the surface normal, thus, preserving the energy conservation law.

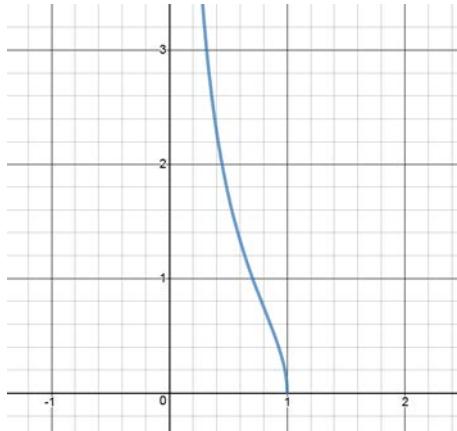


Fig. 2. The plot of cavity height function (vertical axis is the height of theoretical cavity, horizontal axis is Blinn term)

The resulting Normal Distribution Function (NDF) without normalization is:

$$g(\phi, r) = \frac{r}{r + \sqrt{\frac{1}{\cos(\phi)^2} - 1}} \quad (3)$$

where  $f$  is the distribution coefficient,  $r$  is surface roughness, and  $\phi$  is the angle between the surface normal and the cavity normal as well as  $\cos(\phi)$  is the Blinn term.

The integral of this function over  $\phi \in (0; \pi)$  does converge and is equal to:

$$\int_0^\pi \frac{r}{r + \sqrt{\frac{1}{\cos(\phi)^2} - 1}} d\phi = r \frac{\pi r - 2 \log(r)}{2r^2 + 2} \quad (4)$$

The dividing of the proposed distribution function by this integral yields the normalized NDF ready to be used for modelling specular reflections:

$$P(\phi, r) = \frac{2(r^2 + 1)}{(r + \sqrt{\frac{1}{\cos(\phi)^2} - 1})(\pi r - 2 \log(r))} \quad (5)$$

where  $P(\phi, r)$  is the probability of the required cavity,  $r$  is the surface roughness coefficient, and  $\cos(\phi)$  is the Blinn term.

In order to maintain consistency, it is proposed to use the described NDF with a custom geometric masking function based on the same microsurface model.

The proposed geometric masking function describes the decrease of visible highlight due to microsurface cavity being obscured from light source and from the viewer by adjacent microsurface apex as demonstrated in Fig. 3.

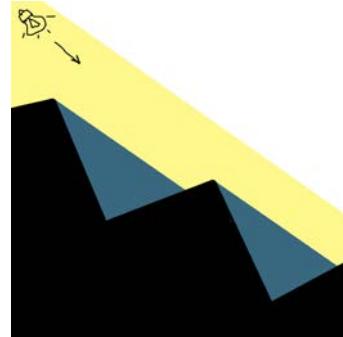


Fig. 3. Demonstration of geometrical masking: surface is black, yellow and blue colors represent light and shadow respectively

The masking term is derived by dividing the length of visible (for view masking) or lit (for light masking) part of microsurface cavity slope by its total length (Fig. 4).

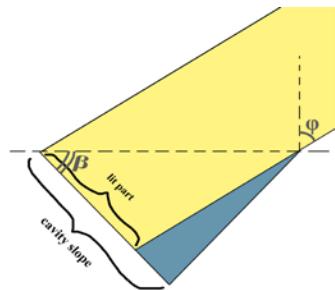


Fig. 4. Masking term derivation ( $\phi$  is the angle between macrosurface normal and light vector,  $\beta$  is the angle between macrosurface and microsurface slope)

The total microsurface slope length can be described as:

$$l_{total} = \frac{1}{2\cos(\phi)} \quad (6)$$

where  $l_{total}$  is the total microsurface slope length and  $\phi$  is the angle between the macrosurface normal and the light vector.

The visible part of the slope is equal to:

$$l_{visible} = \min\left(\frac{1}{\sin\left(\frac{\pi}{2} + \phi - \beta\right)}, l_{total}\right) \quad (7)$$

where  $l_{visible}$  is the visible part of the microsurface slope,  $l_{total}$  is the total microsurface slope length,  $\phi$  is the angle between the macrosurface normal and the light vector, and  $\beta$  is the angle between the microsurface slope and the macrosurface.

Thus, the masking coefficient is equal to:

$$m = \min\left(\frac{2\cos(\phi)}{\cos(\phi - \beta)}, 1\right) \quad (8)$$

where  $m$  is the masking coefficient,  $\phi$  is the angle between the macrosurface normal and the light vector, and  $\beta$  is the angle between the microsurface slope and the macrosurface.

The masking function can be also presented as follows:

$$\frac{2\cos(\phi)}{\cos(\phi - \beta)} = \frac{2\cos(\phi)}{\cos(\phi)\cos(\beta) + \sqrt{(1 - \cos(\phi)^2)(1 - \cos(\beta)^2)}} \quad (9)$$

Thus, the final masking coefficient formula is:

$$m = \min\left(\frac{2\cos(\phi)}{\cos(\phi)\cos(\beta) + \sqrt{(1 - \cos(\phi)^2)(1 - \cos(\beta)^2)}}, 1\right) \quad (10)$$



Fig. 5. Demonstration of highlights: a – measured data from chrome sample, b – proposed model, c – Trowbridge-Reitz model, d – Beckmann model [3]

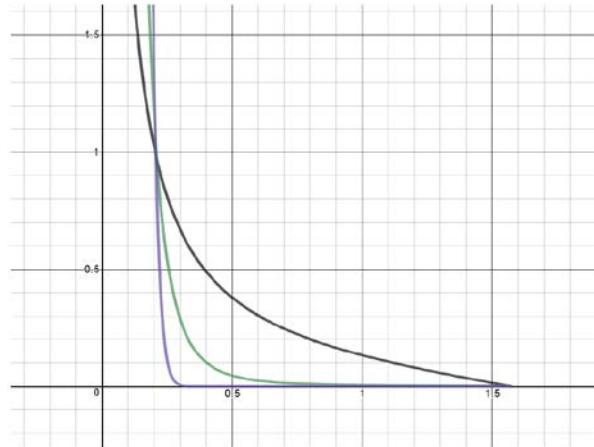


Fig. 6. Comparison of three normal distribution models: the purple line corresponds to Blinn-Phong model, the green line corresponds to Trowbridge-Reitz model, and the black line corresponds to a proposed model

where  $m$  is the masking coefficient,  $\phi$  is the angle between the macrosurface normal and the light vector, and  $\beta$  is the angle between the microsurface slope and the macrosurface.

The total masking coefficient should be calculated as the lowest value between light masking and view masking:

$$m = \min(1, \min\left(\frac{2\cos(\gamma)}{\cos(\gamma)\cos(\beta) + \sqrt{(1 - \cos(\gamma)^2)(1 - \cos(\beta)^2)}}, \frac{2\cos(\phi)}{\cos(\phi)\cos(\beta) + \sqrt{(1 - \cos(\phi)^2)(1 - \cos(\beta)^2)}}\right)) \quad (11)$$

where  $m$  is the total masking coefficient,  $\gamma$  is the angle between the macrosurface normal and the view vector,  $\phi$  is the angle between the macrosurface normal and the light vector, and  $\beta$  is the angle between the microsurface slope and the macrosurface.

#### IV. RESULTS AND DISCUSSION

The proposed model creates a long highlight tail which looks more similar to empirical measurements as demonstrated in Fig. 5 and Fig. 6.

In fact, the highlight tail of the proposed model is brighter than empirical data and, thus, may require further adjustment to achieve better realism in renders.

Using proposed model for rendering specular highlights creates a smoother and more realistic result as shown in Fig. 7.

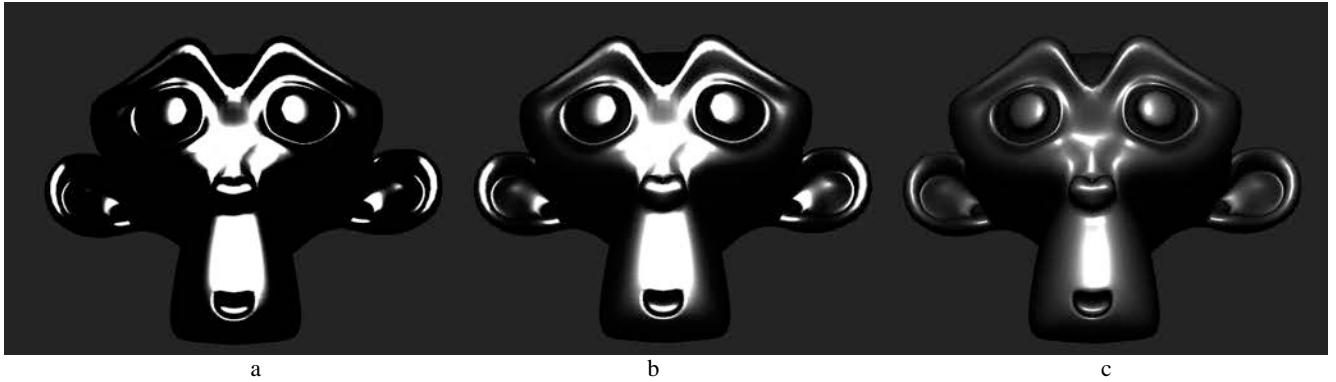


Fig. 7. Demonstration of different normal distribution models: a – Beckmann model, b – Trowbridge-Reits model, c – the proposed model

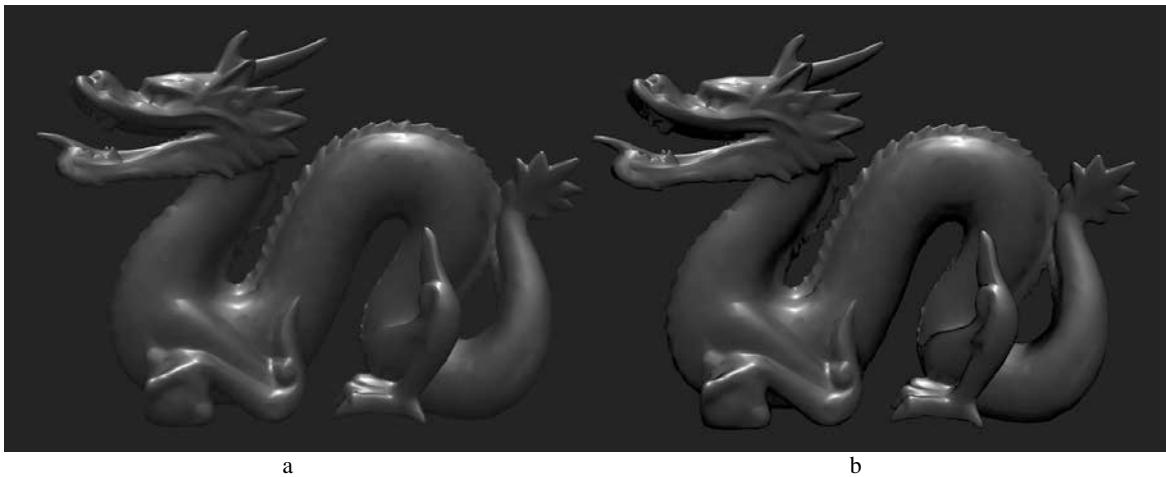


Fig. 8. Proposed masking term demonstration: a – render without a masking term, b – render with the proposed masking term

## V. CONCLUSION

The proposed model retains a long highlight tail even with roughness coefficient extremely close to zero yielding more appealing renders. However, it is also computationally intensive. Besides, it must be noted that the highlight tail of the proposed model in fact seems brighter than measured data and, thus, it may need to be adjusted in order to achieve better realism.

## REFERENCES

- [1] Robert L. Cook, Kenneth E. Torrance, “A reflectance model for computer graphics”, *ACM Transactions on Graphics*, Vol. 1, No. 1, January 1982.
- [2] T. S. Trowbridge and K. P. Reitz, “Average irregularity representation of a rough surface for ray reflection”, *Journal of the Optical Society of America*, Vol. 65, Issue 5, pp. 531-536, 1975.
- [3] P. Beckmann and A. Spizzichino, “The scattering of electromagnetic waves from rough surfaces”, *MacMillan*, 1963.
- [4] Jonathan Dupuy, Eric Heitz, Jean-Claude Iehl, Pierre Poulin, Fabrice Neyret, Victor Ostromoukhov, “Linear Efficient Antialiased Displacement and Reflectance Mapping”, *ACM Transactions on Graphics*, Vol. 32, No. 6, Article 211, pp. 1-11, 2013.
- [5] M.G. Gartley, S.D. Brown and J.R. Schott, “Micro-scale Surface and Contaminate Modeling for Polarimetric Signature Prediction”, *Proceedings of SPIE’08*, Vol. 6972, pp. 1-11, 2008.
- [6] E. Heitz and E. d’Eon, “Importance Sampling Microfacet-Based BSDFs using the Distribution of Visible Normals”, *Proceedings of Eurographics Symposium on Rendering*, Vol. 33 (2014), No 4, pp. 103-112, 2014.
- [7] Eric Heitz, “Understanding the Masking-Shadowing Function in Microfacet-Based BRDFs”, *Journal of Computer Graphics Techniques*, Williams College, Vol. 3 (2), pp.32-91, 2014.
- [8] Eric Heitz, Johannes Hanika, Eugene d’Eon, Carsten Dachsbacher, “Multiple-Scattering Microfacet BSDFs with the Smith Model”, *ACM Trans. Graph.*, Vol. 35, Issue 4, Article 58, pp. 1-14, 2016.
- [9] Bruce Walter, Stephen R. Marschner, Hongsong Li, Kenneth E. Torrance, “Microfacet models for refraction through rough surfaces”, *Proceeding of the 18th Eurographics conference on Rendering Techniques*, pp. 195-206, 2007.
- [10] Eric Heitz , Jonathan Dupuy, Cyril Crassin, Carsten Dachsbaucher, “The SGX microflake distribution”, *Journal ACM Transactions on Graphics*, Vol. 34 Issue 4, Article No. 48, 2015.

# Inverse Dynamic Models in Chaotic Systems Identification and Control Problems

Leonid Lyubchyk<sup>1</sup>, Galyna Grinberg<sup>2</sup>

1. Department of Computer Mathematics and Data Analisys, National Technical University “Kharkiv Polytechnic Institute”, UKRAINE, Kharkiv, 2 Kirpichova str, email: [lyubchik.leonid@gmail.com](mailto:lyubchik.leonid@gmail.com)
2. Department of Economic Cybernetics and Management, National Technical University “Kharkiv Polytechnic Institute”, UKRAINE, Kharkiv, 2 Kirpichova str, email: [glngrinberg@gmail.com](mailto:glngrinberg@gmail.com)

**Abstract:** Inverse dynamic models approach for chaotic system synchronization in the presence of uncertain parameters is considered. The problem is identifying and compensating unknown state-dependent parametric disturbance describing an unmodelled dynamics that generates chaotic motion. Based on the method of inverse model control, disturbance observers and compensators are synthesized. A control law is proposed that ensures the stabilization of chaotic system movement along master reference trajectory. The results of computational simulation of controlled Rössler attractor synchronization are also presented.

**Keywords:** chaotic system, synchronization, disturbance, identification, inverse model, unknown-input observer.

## I. INTRODUCTION

Controlled systems and processes with chaotic dynamics are a matter of unflagging interest in modern control theory and practice [1, 2]. The problem of synchronization of chaotic systems is intensively studied; in this case, control law is designed in such a way that the controlled variables of the slave system follow the reference output of the master system or nonlinear oscillating system stabilized along given reference trajectory in the presence of uncertainties and external disturbances [3, 4].

A typical model of a chaotic system is a linear system with additional nonlinear components dependent on the state, the presence of which determines the appearance of chaotic regimes [5]. Because the system nonlinearity may be treated as a parametric disturbance of nominal model, chaos synchronization problem may be reduced to the disturbance rejection problem, namely, unknown and unmeasurable disturbances eliminating from the systems output along with reference signal tracking.

Recently a number of model-based control methods have been developed for disturbance rejection taking into account the requirements of accuracy, dynamic performance, stability and robustness [6, 7]. In this paper the *inverse model control* approach [8] is applied for chaotic systems synchronization. Inverse models are used for both parametric disturbance identification and compensation, which made it possible to synthesize disturbance decoupling controller, ensure reference signal tracking.

The proposed approach was studied through computational modeling using the example of a controlled

Rössler attractor with signal and parametric disturbances.

## II. PROBLEM STATEMENT

Consider a state-space model of controlled chaotic system with a distinguished nonlinear component, which causes the emergence of chaotic dynamics and interpreted as an uncertain parametric disturbance

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t) + Nf(x(t), \delta), \\ y_c(t) &= Cx(t), \quad y_m(t) = Mx(t), \end{aligned} \quad (1)$$

where  $x(t) \in \mathbb{R}^n$  – chaotic system state vector,  $u(t) \in \mathbb{R}^m$  – control variables vector,  $f(x(t), \delta) \in \mathbb{R}^q$  – state-dependent parametric disturbance with uncertain parameters  $\delta$ ,  $y_c(t) \in \mathbb{R}^r$ ,  $y_m(t) \in \mathbb{R}^p$  – output controlled and measured variables respectively.

Disturbance  $f(x(t), \delta) \in \mathbb{R}^q$  may be treated as unknown input signal for system (1).

Matrices  $S_{CB}(\alpha_1) = CA^{\alpha_1-1}B$ ,  $S_{MN}(\alpha_2) = MA^{\alpha_2-1}N$  are known as Markov parameters of system (1).

Without loss of generality, for simplicity reason, we will assume that  $\text{rank } S_{CB} = m$ ,  $\text{rank } S_{MN} = m$ , where

$$S_{CB} = S_{CB}(1), \quad S_{MN} = S_{MN}(1).$$

Consider two main inverse model problems:

- *Chaotic system identification*, namely, obtaining unknown parametric disturbance estimate  $\hat{f}(t)$  using available measurements  $y_m(t)$  and known control signal  $u(t)$ ;
- *Chaotic system control*, namely, control law  $u(y(t), y^*(t), \hat{f}(t))$  design, which ensure control goal achieving

$$\overline{\lim} \|e_c(t)\|^2 \leq \varepsilon^*, \quad t \rightarrow \infty. \quad (2)$$

where  $e_c(t) = y^*(t) - y_c(t)$  – control error,  $y^*(t)$  – set-point signal given by the reference model

$$\dot{y}^*(t) = A^* \cdot y^*(t) + y_{ref}(t), \quad (3)$$

$\varepsilon^*$  – some sufficiently small constant.

In the chaos synchronization problem, reference model (3) can be considered as a master system [3].

Dynamic system with state vector  $\bar{x}(t) \in \mathbb{R}^{n-q}$

$$\begin{aligned}\dot{\bar{x}}(t) &= A^I \bar{x}(t) + B^I u(t) + B_1^I y(t) + B_2^I \dot{y}(t), \\ \hat{f}(t) &= C^I \bar{x}_k + D^I u(t) + D_1^I y(t) + D_2^I \dot{y}(t),\end{aligned}\quad (4)$$

will be referred to as *inverse dynamic model* of system (1), if the following conditions take place:  $\|\bar{x}(t) - Rx(t)\|^2 \rightarrow 0$ ,  $\|\hat{f}(t) - f(t)\|^2 \rightarrow 0$ , if  $t \rightarrow \infty$ , where  $R_{n-q \times n}$  – some aggregate matrix.

Then  $\hat{f}(t)$  may be treated as unknown input signal  $f(t)$  dynamic estimate, obtained by inverse model (2).

### III. INVERSE DYNAMIC MODEL DESIGN

Let  $z(t) = Rx(t) \in \mathbb{R}^{n-p}$  be aggregated auxiliary variables, where  $R$  is some aggregate matrix, so that  $\text{rank}(M^T | R^T) = n$ .

Take state vector estimate in the form

$$\hat{x}(t) = P \cdot y_m(t) + Q \cdot \bar{x}(t), \quad (5)$$

where matrices  $P \in \mathbb{R}^{n \times p}$ ,  $Q \in \mathbb{R}^{n \times n-p}$  are such that

$$\begin{aligned}MP &= I_p, \quad RQ = I_{n-p}, \quad PM + QR = I_n, \\ MQ &= 0_{p,n-p}, \quad RP = 0_{n-p,p}.\end{aligned}\quad (6)$$

We obtain the aggregated vector  $z(t)$  estimate  $\bar{x}(t)$  by minimal-order unknown-input observer (UIO) [9]:

$$\dot{\bar{x}}(t) = \bar{F}\bar{x}(t) + \bar{G}_1 y_m(t) + \bar{H}\dot{y}_m(t) + \bar{G}_0 u(t). \quad (7)$$

The UIO (7) parameters are determined from disturbance estimate invariance conditions [9, 10]

$$\begin{aligned}(R - \bar{H}M)A - \bar{F}(R - \bar{H}M) &= \bar{G}M, \\ RN - \bar{H}MN &= 0, \quad \bar{G}_0 - RB = 0, \quad \bar{G}_1 = \bar{G} - \bar{F}\bar{H}.\end{aligned}\quad (8)$$

A solution of linear matrix equations (8) are obtained as

$$\begin{aligned}\bar{F} &= R\Pi_N A Q, \quad \bar{G}_0 = RB, \\ \bar{G}_1 &= R\Pi_N A P, \quad \bar{H} = RNS_{MN}^+, \\ \Pi_N &= I_n - BS_{MN}^+ M,\end{aligned}\quad (9)$$

Taking the unknown disturbance estimate as

$$\hat{f}(t) = N^+ (\dot{\hat{x}}(t) - A\hat{x}(t) - Bu(t)), \quad (10)$$

we can obtained the minimal-order state and disturbance observer in the form of system (1) inverse model [10]:

$$\begin{aligned}\dot{\bar{x}}(t) &= R\Pi_N A Q \cdot \bar{x}(t) + R\Pi_N A P \cdot y_m(t) + \\ &\quad + RNS_{MN}^+ \cdot \dot{y}_m(t) + R\Pi_N B \cdot u(t), \\ \hat{x}(t) &= P \cdot y_m(t) + Q \cdot \bar{x}(t), \\ \hat{f}(t) &= \bar{C}_N [\dot{y}_m(t) - MAQ \cdot \bar{x}(t) - \\ &\quad - MAP \cdot y_m(t) - S_{MB}u(t)],\end{aligned}\quad (11)$$

where  $\Pi_N = I_n - NS_{MN}^+ M$ ,  $\Omega_N = I_p - S_{MN}S_{MN}^+$ ,

$$C_N = S_{MN}^+ + N^+ P Q_N.$$

From (1), (11) it follows, that estimate errors vectors  $e_x(t) = x(t) - \hat{x}(t)$ ,  $e_f(t) = f(x(t), t) - \hat{f}(t)$  are given by the equations:

$$\begin{aligned}\dot{\bar{e}}_x(t) &= \bar{F}(R) \cdot \bar{e}_x(t), \\ e_x(t) &= Q \cdot \bar{e}_x(t) \\ e_f(t) &= -C_N MAQ \cdot \bar{e}_x(t).\end{aligned}\quad (12)$$

### III. INVERSE MODEL-BASED CONTROLLER DESIGN

The disturbance rejection control law will be constructed as a function of reference signal and disturbance estimate:

$$\begin{aligned}u^*(t) &= S_{CB}^{-1} \cdot [y_{ref}(t) + C_A \hat{x}(t) - S_{CN} \hat{f}(t)], \\ C_A &= A^* C - CA.\end{aligned}\quad (13)$$

If system structure non-singularity condition takes place

$$\text{rank } \bar{S} = m + q, \quad \bar{S} = \begin{pmatrix} I_m & S_{CB}^{-1} S_{CN} \\ C_N S_{MB} & I_q \end{pmatrix} \quad (14)$$

or equivalently

$$\det \Phi \neq 0, \quad \Phi = I_q - C_N S_{MB} S_{CB}^{-1} S_{CN}, \quad (15)$$

disturbance estimate may be eliminated from the controller equations, which is therefore be regarded as disturbance decoupling controller.

In reality, situations often arise when conditions (14), (15) are not met. In such a case the realizable control law may be obtained using the disturbance estimates, dynamically transformed by the internal auxiliary "fast" filter with small parameters.

As a result, realizable controller are designed by including in its structure an additional internal low-pass filter with small time constant [11]:

$$\begin{aligned}u^*(t) &= S_{CB}^{-1} \cdot [y^*(t) + C_A \hat{x}(t) - S_{CN} \tilde{f}(t)], \\ \tilde{f}(t) &= -\tilde{f}(t) + (1 - \mu) \cdot \hat{f}(t),\end{aligned}\quad (16)$$

where  $0 < \varepsilon \ll 1$ ,  $0 < \mu \ll 1$  - small filter parameters.

From (13), (16) follows, that disturbance compensator equation with internal additional filter take the form:

$$\begin{aligned} \varepsilon \dot{\tilde{u}}(t) &= -\mu \tilde{u}(t) + (1-\mu) \cdot [\varphi_1(t) + \\ &\quad + S_{CB}^{-1} S_{CN} \varphi_2(t)], \\ u^*(t) &= \tilde{u}(t) + \varphi_1(t), \\ \varphi_1(t) &= S_{CB}^{-1} \cdot [y_{ref}(t) + C_A \hat{x}(t)], \\ \varphi_2(t) &= C_N \cdot [\dot{y}_m(t) - MAQ \cdot \bar{x}(t) - MAP \cdot y_m(t)]. \end{aligned} \quad (17)$$

### III. CHAOTIC SYSTEM INVERSE MODEL CONTROL

As an example of proposed approach consider inverse model control of the Rössler attractor under uncertainties:

$$\begin{aligned} \dot{x}_1(t) &= -x_2(t) - x_3(t), \\ \dot{x}_2(t) &= x_1(t) + ax_2(t) + u_1(t) + f_1(t), \\ \dot{x}_3(t) &= -cx_3(t) + u_2(t) + f_1(t) + f_2(x_1(t), x_3(t)), \end{aligned} \quad (18)$$

where

$f_1(t) = \delta_f$ ,  $f_2(x_1(t), x_3(t)) = \delta_c x_3(t) + (1 + \delta_x) x_1(t) x_3(t)$ , are input and parametric disturbances respectively with  $\delta_f, \delta_c, \delta_x$  uncertain parameters.

Using the measurements  $y_1(t) = x_1(t)$ ,  $y_2(t) = x_3(t)$  find the control so the controlled output  $y_c(t) = x_1(t)$  will track set-point signal  $y^*(t)$ , generated by reference model

$$\ddot{y}^*(t) + \alpha_1 \dot{y}^*(t) + \alpha_0 y^*(t) = y_{ref}(t). \quad (19)$$

The control law, which ensures attractor synchronization with reference model, is the following:

$$\begin{aligned} u_2(t) &= (\alpha_0 - 1) \cdot \hat{x}_1(t) + (k - a - \alpha_1) \cdot \hat{x}_2(t) + \\ &\quad + (c - \alpha_1) \cdot \hat{x}_3(t) - 2\hat{f}_1(t) - \tilde{f}_2(t) - y_{ref}(t), \\ u_1(t) &= -k\hat{x}_2(t), \\ \varepsilon \dot{\tilde{f}}(t) &= -\tilde{f}(t) + (1 - \mu) \cdot \hat{f}_2(t), \end{aligned} \quad (20)$$

The state estimates for system (18), obtained by reduced-order UIO, are:

$$\begin{aligned} \dot{\bar{x}}_1(t) &= \rho_1 \bar{x}_1(t) + \bar{x}_2(t) + \\ &\quad + (1 + \pi_1 \rho_1 + \pi_2) \cdot y_1(t) + \pi_1 y_2(t), \\ \dot{\bar{x}}_2(t) &= \pi_2 \bar{x}_1(t) + \pi_1 \pi_2 y_1(t) + \pi_2 y_2(t), \\ \hat{x}_1(t) &= y_1(t), \\ \hat{x}_1(t) &= \bar{x}_1(t) + \pi_1 y_1(t), \quad \hat{x}_3(t) = y_2(t), \end{aligned} \quad (21)$$

where  $\rho_1 = (\pi_1 + a - k)$ ,  $\pi_1, \pi_2$  are tuning parameters.

Corresponding disturbance estimates are

$$\begin{aligned} \hat{f}_1(t) &= \bar{x}_2(t) + \pi_2 y_1(t), \\ \hat{f}_2(t) &= \dot{y}_2(t) + c y_2(t) - \bar{x}_2(t) - \pi_2 y_1(t) - u_2(t). \end{aligned} \quad (22)$$

As a result disturbance decoupling controller with internal filter equation is obtained in the form:

$$\begin{aligned} \varepsilon \dot{\bar{u}}(t) &= v_1 \bar{x}_1(t) - \bar{x}_2(t) + (\zeta_1 - \pi_2) y_1(t) - \alpha_1 y_2(t), \\ u_2(t) &= \bar{u}(t) + v_1 \bar{x}_1(t) - 2\bar{x}_2(t) + \\ &\quad + (\zeta_1 - 2\pi_2) \cdot y_1(t) + (c - \alpha_1 - \varepsilon^{-1}) \cdot y_2(t), \\ \zeta_1 &= \alpha_1 + v_1 \pi_1 - 1, \quad v_1 = k - \alpha_1 - a \end{aligned} \quad (23)$$

Proposed disturbance observer and decoupling controller are investigated by computational simulation.

Simulation results for Rössler attractor model parameters  $a = 0.2$ ,  $c = -5.7$ , observer and controller parameters  $\pi_1 = -1$ ,  $\pi_2 = -2$ ,  $\varepsilon = 0.01$ ,  $\mu = 0$ ,  $k = 2.2$ , and reference model parameters  $\alpha_0 = 5$ ,  $\alpha_1 = 6$  are presented below.

Disturbance  $f_1(t)$  was modeled input signal disturbance as a step wave function, reference model input signal  $y_{ref}(t)$  adopted in the form of harmonic function.

At Fig.1, 2 the state variables and phase plane of controlled Rössler disturbed attractor are presented.

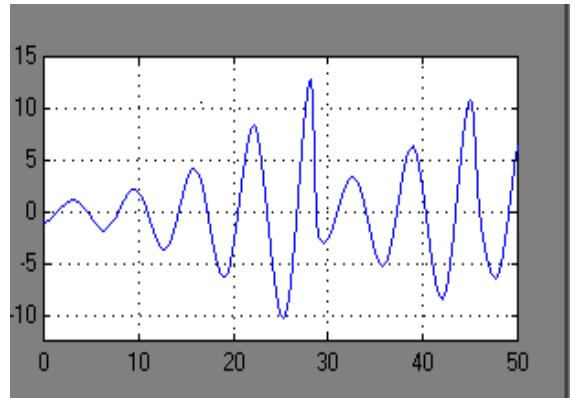


Fig.1. Dynamics of the disturbed attractor.  
State variable  $y_1(t)$

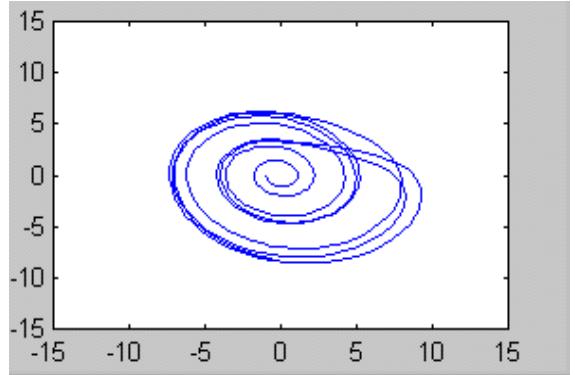


Fig.2. Dynamics of the disturbed attractor.  
Phase plane  $(y_1(t), y_2(t))$

Disturbances estimations obtained by (21), (22) are depicted in Fig. 3, 4 and control and output variables obtained in accordance the control law (20), (23) are presented at Fig. 5, 6.

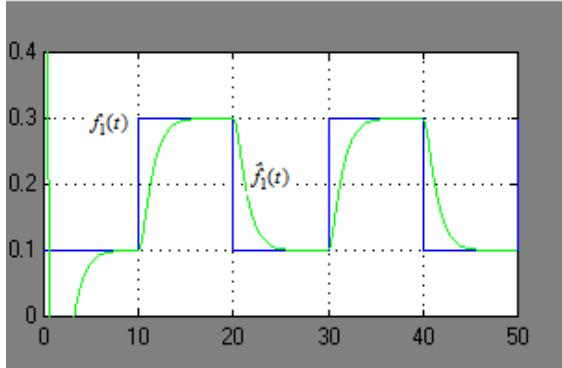


Fig.3. Disturbances estimation  $f_1(t), \hat{f}_1(t)$  in open-loop system

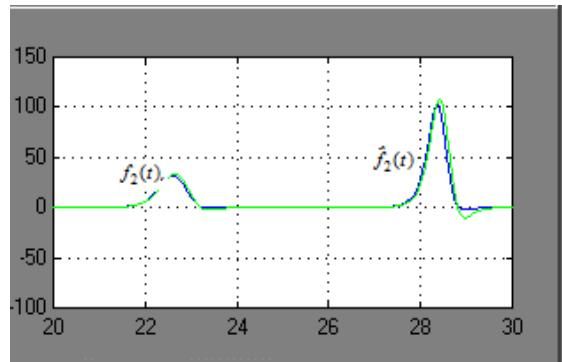


Fig.4. Disturbances estimation  $f_2(t), \hat{f}_2(t)$  in open-loop system

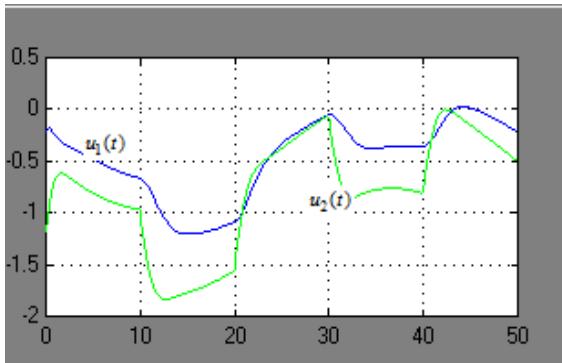


Fig.5. Control variables  $u_1(t), u_2(t)$

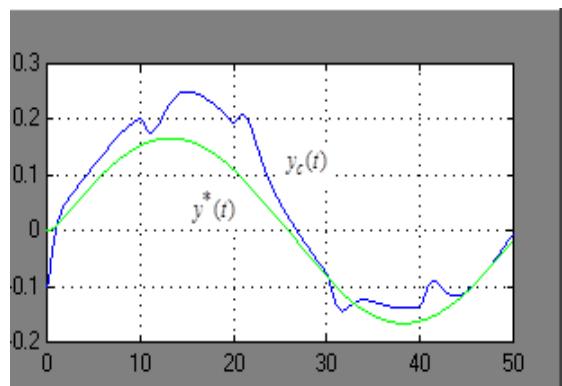


Fig.6. Set-point signal  $y^*(t)$  and output variables  $y_c(t)$

Simulation results for chaotic system synchronization problem demonstrated high accuracy of disturbances decoupling for broad range of parameters deviation.

#### IV. CONCLUSION

In this paper we presented inverse model-based approach to chaotic systems synchronization problem. The proposed method allows us to decompose the problem into the stage of structural synthesis of inverse models and their parametric synthesis or optimization. This significant advantage of the method of inverse dynamic models is the possibility of real-time reconstruction of signals of complex shape in the absence of information on their structure. This, in turn, makes it possible to efficiently solve problems of compensation of nonlinear state-dependent, which makes it possible to suppress sources of chaotic dynamics and simplifies the solution of synchronization problems. Thus proposed approach seems to be quite universal and can be used to solve various problems of controlling chaotic systems.

The implementation of the proposed control requires differentiating the measured output signals in real time, for which differentiators based on sliding modes can be used. Further development of the proposed approach is associated with the development of robust methods for inverse models design under conditions of uncertain deviations of the parameters of the chaotic object model.

#### REFERENCES

- [1] G. Chen, (Ed), "Controlling Chaos and Bifurcations in Engineering Systems", CRC Press, Boca Raton, USA, 1999.
- [2] Fradkov, A. Pogromsky, "Introduction to control of oscillations and chaos", World Scientific, Singapore, 1998.
- [3] Ju Park, "Chaos synchronization of a chaotic system via nonlinear control", *Chaos, Solitons and Fractals*, no. 25, pp. 579–584, 2005.
- [4] S. Mohammadpour, T. Binazadeh, "Robust finite-time synchronization of uncertain chaotic systems", *Systems Science & Control Engineering*, vol. 6, 2018.
- [5] Z. Shen, J. Xiong, and Y. Wu, "Uncertain Unified Chaotic Systems Control with Input Nonlinearity via Sliding Mode Control", *Mathematical Problems in Engineering*, Hindawi, vol. 2016, 9 p., 2016.
- [6] Wolovich W.A., "Automatic Control Systems: Basic Analysis and Design", Philadelphia, PA: Saunders, 1995.
- [7] Ya. Tsyplkin, U. Holmber, "Robust stochastic control and internal model control", *Intern. Journal of Control*, vol. 61, no 4, pp. 809-822, 1995.
- [8] L. Lyubchyk, "Disturbance rejection in linear discrete multivariable systems: inverse model approach", *IFAC Proceedings Volumes*, 2011.
- [9] M. Hou, P. Muller, "Design of observers for linear systems wits unknown inputs", *IEEE Trans. on Automatic Control*, vol.37, pp.871-875, 1992.
- [10] L. Lyubchyk, "Inverse model control and sub-invariance in linear discrete multivariable systems", in *Proc. of 3-rd European Control Conference*, Roma, vol. 4, part 2, pp. 3651-3659, 1995.
- [11] J. Lunze, "Robust Multivariable Feedback Control", Prentice Hall, 1988.

# Adjustment of the Model of the Agent-Determinant Type in the Forecasting of Pollution on the Section of the City Road

Mykhailo Susla<sup>1</sup>, Roman Pasichnyk<sup>2</sup>, Natalia Pasichnyk<sup>3</sup>, Andriy Melnyk<sup>4</sup>

Department of Computer Science, Ternopil National Economic University, UKRAINE, Ternopil, 8 Chelhova str., email: [suslamisha91@gmail.com](mailto:suslamisha91@gmail.com)<sup>1</sup>, [roman.pasichnyk@gmail.com](mailto:roman.pasichnyk@gmail.com)<sup>2</sup>, [natalia.pasichnyk@gmail.com](mailto:natalia.pasichnyk@gmail.com)<sup>3</sup>, [melnyk.andriy@gmail.com](mailto:melnyk.andriy@gmail.com)<sup>4</sup>.

**Abstract:** The problem of constructing a dynamic model of pollution is considered. The model of pollution in the form of a set of differential equations is proposed, which is identified by means of difference expression with subsequent refinement using gradient methods. Numerical experiments allow you to choose the model that best approximates the experimental data.

**Keywords:** air pollution; mathematical model; set of ordinary differential equations; difference expressions; Levenberg–Marquardt method.

## I. INTRODUCTION

The rapid growth of the number of vehicles brings out the problem of control over air pollution with motor vehicles. The latter affects the level of respiratory and other diseases for people living in the areas of high traffic density. Therefore, there is a need for sufficiently accurate monitoring of information on the level of air pollution to make managerial decisions regarding the configuration of residential quarters, design and reconstruction of roads.

The distribution of pollution in the city is characterized by significant spatial heterogeneity, since emissions to the atmosphere are carried out from the network of roads, and the pollution level declines rapidly as we move away from the pollution source [1, 2]. These features take into account the Land use regression (LUR) method, which combines measurements of air pollution in a relatively small number of locations characterized by qualitatively different types of pollution, and the construction of statistical models based on measurements taking into account the features of the points of observation.

Numerous researchers use LUR to estimate the concentration of contaminants in a number of cities in Canada, the United States and Europe [3]. However, this method allows you to build only stationary models. At the same time, it is necessary to build dynamic models for deeper understanding of the pollution effects using apparatus differential equations.

Such a model, which takes into account the influence of several key factors, should be as simple as possible and at the same time sufficiently precise. If we allow interaction of factors with each other, in the simplest way it is modeled using the product of the corresponding indicators.

The generalization of this approach is a model-type agent-determinant that reflects the evolution of the agent under the

influence of the determinant so that a significant concentration of the determinant does not compensate for the low concentration of the agent. A representative of models of this type is the Monod model. In a number of simulated situations, such a generalized model is extremely effective [8]. This work is devoted to the study of the possibility of using models of the specified type in the simulation of pollution concentration.

## II. MODEL OF POLLUTION ON THE LOCAL SECTION OF AN URBAN ROAD

In order to construct a model of the dynamics of pollution in an area where it is potentially high, we have to identify the main variables that affect it. No significant influence of humidity and temperature on the dynamics of pollution levels  $X$  has been found after the analysis of the measurements obtained with the help of special sensors. Instead, a significant effect of the traffic intensity  $R$  and wind speed  $V$  has been established.

The apparatus of differential equations is chosen to simulate the dynamics of pollution, since it's much more flexible than regression relations. Numerical differentiation is very sensitive to random perturbations in the measurement results, so it is subjected to multiple smoothing by the method of moving average.

The criterion of multiplicity of smoothing served The minimization of the correlation between the remnants of measurements after their elimination from the smoothed values served as a criterion of multiplicity of smoothing.

When constructing the differential equation of the dynamics of pollution it is taken into account that the growth of pollution is associated with an increase in the intensity of pollutants, ie, the movement of vehicles. Contamination reduce occurs as a result of their dispersal, which is associated with the speed of the wind.

However, the speed of diffusion of contaminants depends not only on the mentioned factor but also on the product of it and of the concentration of contaminants themselves. Since pollution itself decomposes over time, the contaminants concentration decreases in proportion to the pollution itself. The above statements are established on the basis of data analysis and confirmed during the preceding procedures of parametric identification. As a result, we come to the following differential equation of the dynamics of pollution on the road section

$$\frac{dX(t)}{dt} = p_1 R(t) - (p_2 + p_3 V(t)) X(t), \quad (1)$$

$$X(t_0) = X_0. \quad (2)$$

where  $X$  is pollution concentration;  $R$  is traffic intensity;  $V$  is wind speed;  $p_1, \dots, p_4$  are model parameters.

In an empirically constructed model, the interaction of contaminants is described by their product with a constant relative intensity of interaction. In some cases, the actual intensity of the interaction may vary with the change in the determinant characteristic, in this case, the wind speed. Often there is a variable intensity of interaction, which is lower at small values of the determinant and is obtained at the maximum value with saturation at large values of the determinant. This intensity is fed by a multiplicand of the Monod type

$$\frac{V(t)}{p_4 + V(t)} \quad (3)$$

With its application, the model equation (1) takes the form

$$\frac{dX(t)}{dt} = p_1 R(t) - p_2 X(t) - p_3 V(t) X(t) \frac{V(t)}{p_4 + V(t)} \quad (4)$$

As a result, we obtain a more complex differential equation containing an additional parameter, which is included nonlinearly. To compare the effectiveness of the proposed models, they need to be identified.

### III. THE IDENTIFICATION METHODS OF POLLUTION MODEL

It is necessary to establish a method of parametric identification of the pollution model after we have built it. Usually, identification is carried out by minimizing the appropriate quality functional. One of the simplest quality functional is the square root, which is used in the least squares method.

Since the differential equation is nonlinear, its quality functional has a large number of local extrema. The general approach to building a global extremum of this kind is the use of methods of random search, the method of the directing cone of Rastrigin, in particular. However, this method requires great amount of computing resources and the development of special procedures for locating local extremes to find a global one.

At the same time, taking into account the peculiarities of certain classes of tasks, it is possible to set up search domains containing a single global extremum. In particular, a whole class of methods of this kind is proposed for the identification of models of systems with limiting factors.

In these methods, the initial approximation of the values of the models' parameters is based on the difference ratios and scanning the values of one of the key parameters on the grid. The parameters of this grid are also pre-evaluated. The initial approximation is further specified by the gradient method.

These methods have shown their high efficiency and therefore the corresponding method of identification of systems of nonlinear differential equations modeling the

dynamics of processes with limiting factors is chosen as the basis for the method of model (1) - (2) identification [4].

Since this differential equation (1) is much simpler, the identification method itself is also simplified. At the initial stage, we construct a system of linear equations with respect to the parameters basing on the difference relations:

$$\frac{X_{i+1} - X_{i-1}}{t_{i+1} - t_{i-1}} = p_1 R(t_i) - (p_2 + p_3 V(t_i)) X(t_i) \quad i = i_1, i_2, i_3. \quad (5)$$

Where  $X_i$ ,  $R_i$ ,  $V_i$  are the values of corresponding functions at the moment of time  $t_i$

The ratio for constructing the initial approximations of the coefficients of the differential equation must reflect the most significant features of the resulting function of the process, that is, in this case, the dynamics of pollution. The numbers of the identification points are chosen in case of the maximum absolute values of the derivatives of the pollution concentration function. Further, the initial values of the parameters of the model are specified by the method of least squares.

The procedure for identifying a differential equation (4) is somewhat more complicated. It includes a method for checking the values of a nonlinear parameter  $p_4$  on a grid, whose parameters are selected experimentally. After selecting a specific parameter value  $p_4$  for choosing the initial values of other parameters, an analogue of the system of linear equations (5) is used.:

$$\begin{aligned} \frac{X_{i+1} - X_{i-1}}{t_{i+1} - t_{i-1}} &= p_1 R(t_i) - p_2 X(t_i) - \\ &- p_3 V(t_i) X(t_i) \frac{V(t_i)}{p_4 + V(t_i)} \quad i = i_1, i_2, i_3 \end{aligned} \quad (6)$$

### IV. NUMERIC EXPERIMENTS

Let us demonstrate the possibilities of the proposed methodology on the example of modeling the daily dynamics of pollution on one of the streets of the city of Ternopil. Discrete observations are interpolated using piecewise Hermite interpolation. In particular, the dynamics of wind speed during the day of observation is given on Figure 1.

The following figures show the smoothed results of observations of pollution and traffic. As you can see, the dynamics of the concentration of pollution is quite complicated. To verify the reality of the identification of the proposed model, we analyze the dynamics of the left and elements of the right-hand side of the differential equation (1), given in Figure 4.

It is worth noting that the components of the left parts of the differential equation are brought to comparable values with the pollution derivative by multiplication on corresponding scaling multipliers keeping "+" or "-" sign, how they are included in the equation. The comparison of these functions reveals some similarity in their behavior, as well as the complexity of the task of bringing their sum to zero with the help of just three constant coefficients.

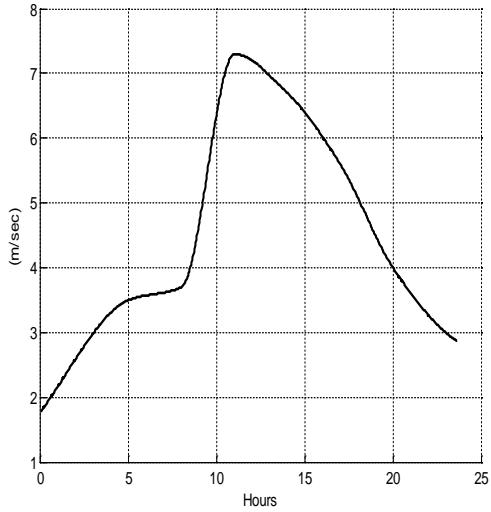


Fig 1. Observation of wind speed during the day

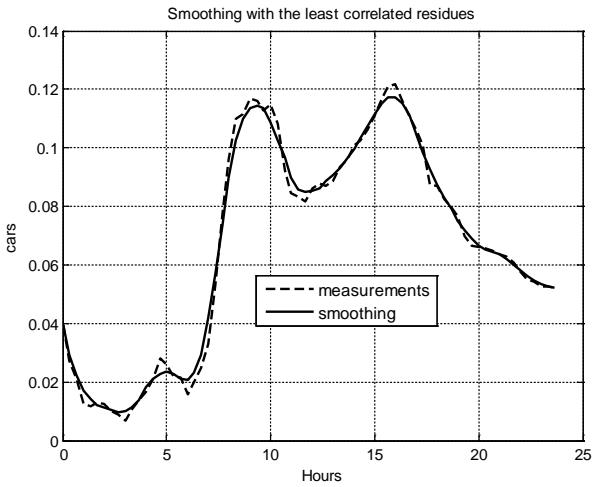


Fig 2. Observation of pollution during the day

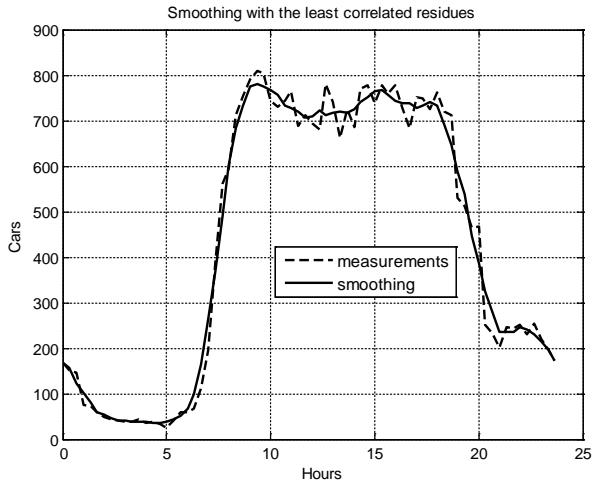


Fig 3. Observation of traffic during the day

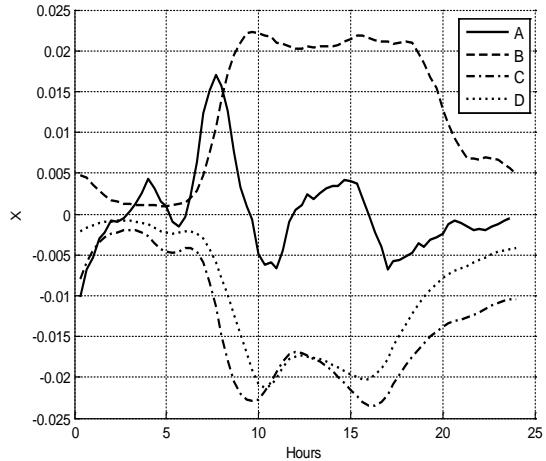


Fig 4. The comparison of the left hand and elements of the right-hand parts of the equation (1), where a) pollution derivative; b) traffic; c) pollution; d) wind speed and pollution product.

By the criterion of the maximum of the derivative module the points 7.3, 10.7, 14.3 have been selected among the internal points of the time interval. As a result of the solution of the system of linear equations (3) the following values of the parameters of the differential equation (1) are obtained:  $p_1 = 7.0958e-4$ ,  $p_2 = 3.0172$ ,  $p_3 = 0.3175$ .

After optimization of the initial approximation of coefficients using the gradient method of Levenberg-Marquardt, only the first coefficient was somewhat specified to the value  $p_1 = 7.0942e-4$ , which allowed to somewhat decrease the average identification error.

The identification results for the equation (1)-(2) are presented on the figure 5, average identification error is 8.6%. Details of the distribution of errors in points of observation can be found using Figure 6.

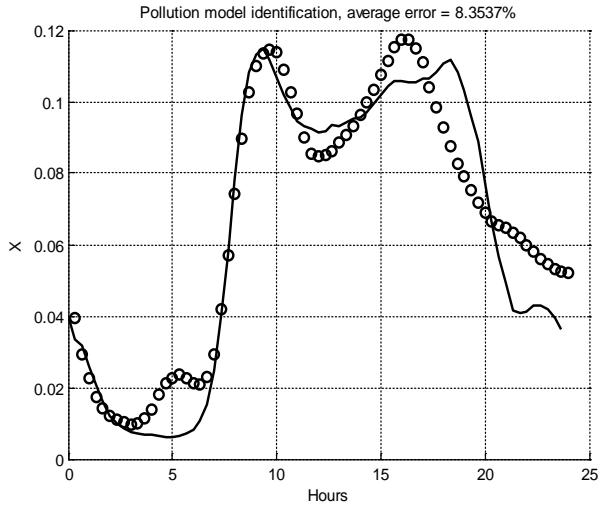


Fig 5. Approximation of observed pollution using identified model (1) - (2)

The question arises whether we can significantly improve the accuracy of identification by using a more complex model (4) - (1)?

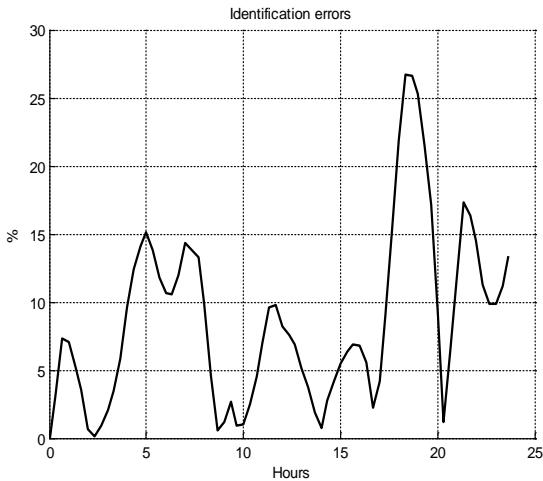


Fig 6. Distribution of model (1)-(2) identification errors

We will investigate this question experimentally. The identification of the model (4) - (1), with the results of which is shown in Figure 7, was carried out by checking the values of the parameter  $p_4$  on the specially selected grids and the identification procedure given by the relations (6). As a result of solving the system of linear levels (6), refining the parameters by the Levenberg-Marquardt method and selecting the parameter  $p_4$  as the basis of the performed calculations, the criterion for minimizing average ratios is the following values of the parameters of the differential equation (1):

$$p_1 = 6.9316_{e-4}, p_2 = 2.9586, p_3 = 0.3683, p_4 = 0.6810$$

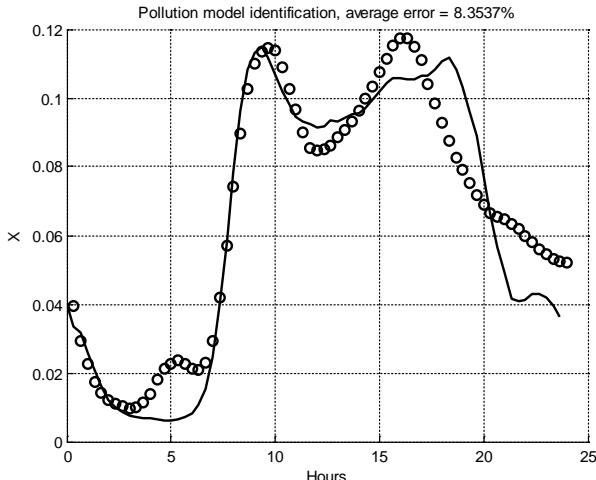


Fig 7. Approximation of observed pollution using identified model (4) - (2)

The average error of identification was 8.35%, which improves insignificantly significantly the result of the approximation of experimental data using the model (1) - (2). The analysis of the distribution of errors of approximation by model (4) - (2) also did not reveal any significant differences with the distributions of errors in the model (1) - (2).

### III. CONCLUSION

As a result of the conducted research, a model of pollution on the local section of the city road was proposed in a form of

an ordinary differential equation, which includes observation of typical daily traffic and daily forecast of wind speed. The methods of parametric identification of the constructed models were proposed. The methods are based on the use of difference approximation of the differential equation in specially selected points for construction of initial approximations of the model coefficients with their further refinement by the Levenberg-Marquardt method. Also, based on the use of selection of the  $p_4$  parameter value included in the equation (4) nonlinearly in the sequence of special grids. As a result of the experimental study, the proposed models established their practical identity with the accuracy of the approximation of experimental data. Therefore, the simplest of them, namely the model (1) - (2), should be preferred.

The chosen model can serve as the basis for constructing a dynamic map of pollution of the city.

### REFERENCES

- [1] K.M.Antropov, J.I.Kazmer, A.N. Varaksin. "Assesing spatial variability of air pollution in industrial city with land use regression (review)". Ecological systems and devices. Moscow, vol.1, pp. 28-41, 2010.
- [2] T. Bush, S. Smith, K. Stevenson, S. Moorcroft, "Validation of nitrogen dioxide diffusion tube methodology in the UK", Atmospheric Environment, vol. 35, 2, 2001.
- [3] I. Voytyuk, M. Dyvak, V. Spilchuk. "Research of quality characteristics of models structure in kind of interval difference operator". Proceedings of International Conference CADSM'2011, Polyana-Svalyava, pp. 87, 2011.
- [4] N. Porplytsya, M. Dyvak,, I. Spivak,, I. Voytyuk, "Mathematical and algorithmic foundations for implementation of the method for structure identification of interval difference operator based on functioning of bee colony", 13th International Conference: The Experience of Designing and Application of CAD Systems in Microelectronics, CADSM 2015, pp. 196-199, 2015.
- [5] N. Ocheretnyuk, I. Voytyuk, M. Dyvak, Ye. Martsenyuk, "Features of structure identification the macromodels for nonstationary fields of air pollutions from vehicles", Modern Problems of Radio Engineering, Telecommunications and Computer Science - Proceedings of the 11th International Conference, TCSET'2012, pp. 444.
- [6] T. Sahsuvaroglu, A. Arain, B. Beckerman, P. Kanaroglou, J.R. Brook, N. Finkelstein, M.M. Finkelstein, N.L.Gilbert, B. Newbold, M.Jerrett. "A LUR model for predicting ambient concentrations of nitrogen dioxide in Hamilton, Canada", J. Air Waste Manage. Assoc., vol. 56, 8, 2006.
- [7] R. Pasichnyk, M. Dyvak, N. Pasichnyk. "Identification and modeling of limiting factors systems". Data Stream Mining & Processing (DSMP), IEEE First International Conference, pp. 336-34, 2016.
- [8] R. Pasichnyk. "Modeling of resources accumulation and operational management in biotechnology, biomedical and web information systems". Computational Problems of Electrical Engineering, vol. 4, 2, pp. 37-46, 2014.

# Modeling of Adsorption and Desorption of Hydrocarbons in Nanoporous Catalyst Zeolite using Nonlinear Langmuir's Isotherm

Mykhaylo Petryk, Dmytro Mykhalyk, Maria Petryk, Igor Boyko, Ivan Mudryk

Department of Software Engineering ,  
Ternopil Ivan Puluj National Technical University, UKRAINE, Ternopil, 56 Ruska str.,  
email: mykhaylo\_petryk@tu.edu.te.ua

*Abstract:* The theoretical bases of mathematical modeling of nonisothermal adsorption and desorption in nanoporous Catalyst Zeolite for nonlinear Langmuir's isotherm are given. An effective linearization scheme for the nonlinear model is realized. High-speed analytical solutions of the system of linearized boundary-value problems of adsorption and desorption in nanoporous media are substantiated and obtained.

**Keywords:** mathematical modeling, adsorption desorption process, nanoporous catalyst zeolite, Langmuir's isotherm.

## I. INTRODUCTION

The quality of mathematical models for adsorption and desorption processes of hydrocarbons in nanoporous catalytic media determines the effectiveness of technological solutions for neutralizing and reducing exhaust emissions of internal combustion engines, the number of which is rapidly increasing, contributing to global warming crisis. [1, 2].

At present, many experimental and theoretical studies of such processes are conducted, especially studies on the improvement of mathematical models, taking into account the influence of various factors that limit the internal kinetics of adsorption and desorption in nanopores of catalytic media. A detailed analysis of these works was made in [3].

This paper describes the theoretical foundations for modeling non-isothermal adsorption and desorption in nanoporous catalysts for a nonlinear isotherm obtained by the American physicist, Nobel Prize winner Erwin Langmuir, who most fully determines the mechanism of adsorption equilibrium for micro- and nanoporous systems of ZSM-5 zeolites.

## II. DESCRIPTION OF KINETIC PROCESSES AND MATHEMATICAL MODEL

A general description of the interaction of a diffusing gas stream in a biporous pore system of a catalytic medium of nanoporous particles, taking into account the main limiting factors of internal kinetics of mass transfer, including the interaction of micro- and macro transfer, is given in [3].

The main hypothesis assumed for the model is adsorption interaction between adsorbent molecules and active adsorption centers on the phase separation surface in micro- and nanopores of crystallites is determined on the basis of

Langmuir's non-linear adsorption equilibrium function taking into account the physical prerequisites [4- 6]:

1. Adsorption is localized and is caused by dispersion forces, the interaction of which is established by Lenard and the electrostatic forces of attraction and repulsion, the mechanism of which is described by Van-Der-Waals [4].
  2. Adsorption takes place in active centers on the surface of adsorbent distributed throughout the internal surface of the micro- and nanopores.
  3. Each active center adsorbs only one molecule of adsorbent and its molecular layer of adsorbate is formed on the surface.
  4. Adsorbed molecules are retained by active centers during a certain time, depending on the temperature.

Proceeding from these, the adsorption equilibrium function (adsorption isotherm) of Langmuir type, which describes the adsorbent phase transition from gas flow to the micro- and nanopores of catalytic medium, will be determined by a nonlinear dependence establishing relationship between equilibrium concentration and adsorption value [5].

$$a \equiv f(c_{eq}) = a_{full} \frac{bc_{eq}}{1+bc_{eq}}. \quad (1)$$

Here  $a_{full}$ ,  $0 < b < 1$  are the empirical coefficients that depend on properties of nanoporous medium and diffused substance:  $a_{full}$  - the concentration (amount) of adsorbate in micro- and nanopores of zeolite with complete filling of the adsorption centers.

The refined kinetics of nonisothermal adsorption and desorption for exhaust gas neutralization systems in nanoporous catalysts, taking into account the nonlinear function of adsorption equilibrium and the given physical justifications, can be described by the following system of nonlinear partial differential equations [5, 6]:

$$\begin{aligned} & \frac{\partial c(t, z)}{\partial t} + \frac{\partial a(t, z)}{\partial t} + u \frac{\partial c}{\partial z} = D_{\text{inter}} \frac{\partial^2 c}{\partial z^2}, \\ & -H \frac{\partial T(t, z)}{\partial t} - u h_g \frac{\partial T}{\partial z} - Q \frac{\partial a}{\partial t} - X^2 T + \Lambda \frac{\partial^2 T}{\partial z^2} = 0, \quad (2) \\ & \frac{\partial a}{\partial t} = \beta \left( c - \frac{1}{b} \frac{a}{a_{\text{full}} - a} \right). \end{aligned}$$

with initial conditions:

### a) adsorption

### b) desorption

$$c(t, z)|_{t=0} = 0,$$

$$T(t, z)|_{t=0} = T_0,$$

and boundary condition:

a) adsorption

$$c(t, z)|_{z=0} = c_{in},$$

$$\frac{\partial}{\partial z} c(t, z)|_{z=\infty} = 0, \quad \frac{\partial}{\partial z} c(t, z)|_{z=0} = 0,$$

$$T(t, z)|_{z=0} = T_{in},$$

$$\frac{\partial}{\partial z} T(t, z)|_{z=\infty} = 0,$$

$$T(t, z)|_{z=0} = T_{in}(t),$$

$$c(t, z)|_{t=0} = c_0,$$
(3)

$$T(t, z)|_{t=0} = T_0,$$
(4)

b) desorption

$$c(t, z)|_{z=0} = c_{in}(t),$$
(5)

$$\frac{\partial}{\partial z} c(t, z)|_{z=\infty} = 0,$$
(6)

$$\frac{\partial}{\partial z} T(t, z)|_{z=\infty} = 0,$$
(7)

$$\frac{\partial}{\partial z} T(t, z)|_{z=0} = 0$$
(8)

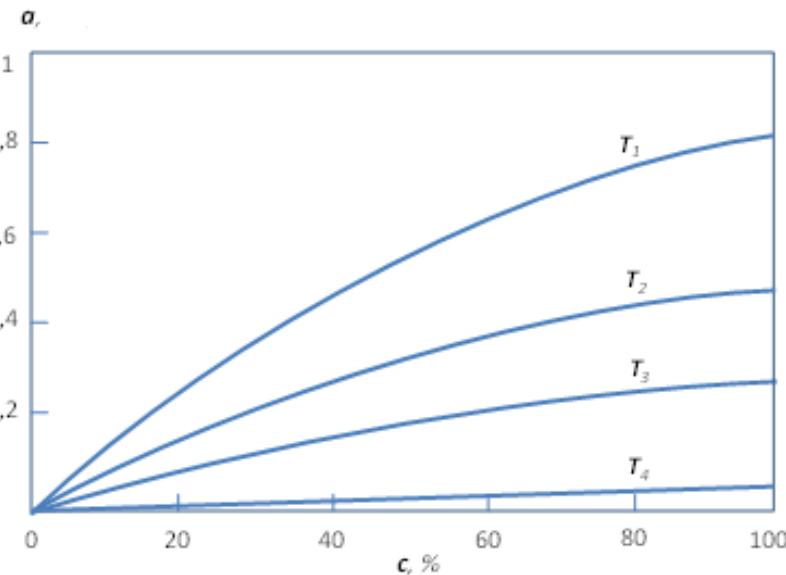


Fig.1 The ZSM-5 zeolite adsorption isotherm in the temperature range from 0 to 350 °C.

As can be seen from fig. 1, adsorption value increases according to the nonlinear law as the adsorbent concentration in the gas phase increases, accompanied by the "filling" of active adsorption centers on the surface of the micropores, and decreases with increasing medium temperature ( $T_1 < T_2 < T_3 < T_4$ ) [5].

Substituting the expanded expression (9) instead of the dependence in the third equation of system (3), we obtain

$$\frac{\partial a}{\partial t} = \beta(c - \gamma a(z, t) - \varepsilon a^2(z, t)) \quad (10)$$

### III. THE LINEARIZATION OF A NONLINEAR MODEL

The problem (2) - (8), taking into account the approximated kinetic equation of phase transformation (10) containing a small parameter  $\varepsilon$ , is a mixed boundary-value problem for a nonlinear system of second-order partial differential equations. The solution of problem (2) - (8) will be obtained using asymptotic expansions in small parameter  $\varepsilon$  in the form of following power series [7, 8]:

$$c(t, z) = c_0(t, z) + \varepsilon c_1(t, z) + \varepsilon^2 c_2(t, z) + \dots,$$

$$T(t, z) = T_0(t, z) + \varepsilon T_1(t, z) + \varepsilon^2 T_2(t, z) + \dots, \quad (11)$$

Taking into account that  $\frac{a}{a_{full}} < 1$ , the Maclaurin's series,

we obtain:

$$c_{eq}(a) \equiv \varphi(a) = \frac{1}{b} \frac{a/a_{full}}{1-a/a_{full}} \approx \gamma a(t, z) + \varepsilon a^2(t, z), \quad (9)$$

where  $\gamma = \frac{1}{ba_{full}}$  is adsorption constant, which describes linear component of the adsorption equilibrium function  $c_{eq}(a)$  (according to Henry's law),  $\varepsilon = \frac{1}{b(a_{full})^2}$  - is a small parameter that takes into account the nonlinear component of the adsorption isotherm.

$$a(t, z) = a_0(t, z) + \varepsilon a_1(t, z) + \varepsilon^2 a_2(t, z) + \dots.$$

As a result of the substitution of asymptotic sums (11) into equations (2) and taking into account (10), the initial nonlinear problem (2) - (8) is parallelized into two types of linear problems [8]:

*The problem  $A_0$  (zero approximation):* to find a solution for the system of partial differential equations:

$$\frac{\partial c_0(t, z)}{\partial t} + \frac{\partial a_0(t, z)}{\partial t} + u \frac{\partial c_0}{\partial z} = D_{inter} \frac{\partial^2 c_0}{\partial z^2}, \quad (12)$$

$$-H \frac{\partial T_0(t, z)}{\partial t} - u h_s \frac{\partial T_0}{\partial z} - Q \frac{\partial a_0}{\partial t} - X^2 T_0 + \Lambda \frac{\partial^2 T_0}{\partial z^2} = 0, \quad (13)$$

$$\frac{\partial a_0}{\partial t} = \beta(c_0 - \gamma a_0), \quad (14)$$

with initial and boundary conditions of initial problem.

*The problem  $A_n$  (n-th approximation with zero initial and boundary conditions):* to find a solution for system of equations:

$$\frac{\partial c_n(t, z)}{\partial t} + \frac{\partial a_n(t, z)}{\partial t} + u \frac{\partial c_n}{\partial z} = D_{inter} \frac{\partial^2 c_n}{\partial z^2}, \quad (15)$$

$$-H \frac{\partial T_n(t, z)}{\partial t} - u h_g \frac{\partial T_n}{\partial z} - Q \frac{\partial a_n}{\partial t} - X^2 T_n + \Lambda \frac{\partial^2 T_n}{\partial z^2} = 0, \quad (16)$$

$$\frac{\partial a_n}{\partial t} = \beta \left( c_n - \gamma a_n - \sum_{i=0}^{n-1} a_i(t, z) a_{n-1-i}(t, z) \right) \quad (17)$$

with zero initial and boundary conditions.

We construct analytic solutions of problems  $A_0$  and  $A_n; n = \overline{1, \infty}$  using the Heaviside's operation method [9, 10].

The problem  $A_0$  is linear concerning to zero approximation  $a_0$ ; The problem  $A_n; n = \overline{1, \infty}$  is linear concerning to the nth approximation  $a_n$  and nonlinear concerning to all previous n-1 approximations. All equations of problems are obtained by linearizing the nonlinear differential equation of the internal adsorption kinetics with asymptotic sums (11), grouping the terms in the left and right sides of the equations and the conditions of the original boundary value problem for equal powers of a small parameter.

Having determined,

$$\begin{aligned} L[c(t, z)] &\equiv c^*(p, z) = \int_0^\infty c(t, z) e^{-pt} dt, \\ L[T(t, z)] &\equiv T^*(p, z) = \int_0^\infty T(t, z) e^{-pt} dt, \\ L[a(t, z)] &\equiv a^*(p, z) = \int_0^\infty a(t, z) e^{-pt} dt \end{aligned} \quad (18)$$

where p is a complex Laplace transform parameter, we obtain in the Laplace images  $A_0^*$  and  $A_n^*$  the above boundary value problems.

*The problem  $A_0^*$*

$$\frac{d^2 c_0^*(p, z)}{dz^2} - u_1 \frac{dc_0^*}{dz} - q_1^2(p) c_0^* = -\mathcal{F}_{c_0^*}(p), \quad (19)$$

$$\frac{d^2}{dz^2} T_0^* - u_2 \frac{d}{dz} T_0^* - q_2^2(p) T_0^* = -\mathcal{F}_{T_0^*}(p), \quad (20)$$

$$a_0^*(p, z) = \beta \frac{1}{p + \beta \gamma} c_0^*(p, z), \quad (21)$$

*The problem  $A_n^*$*

$$\frac{d^2 c_n^*}{dz^2} - u_1 \frac{dc_n^*}{dz} - q_1^2(p) c_n^* = -\mathcal{F}_{c_n^*}(p, z), \quad (22)$$

$$\frac{d^2}{dz^2} T_n^* - u_2 \frac{d}{dz} T_n^* - q_2^2(p) T_n^* = -\mathcal{F}_{T_n^*}(p, z), \quad (23)$$

$$a_n^*(p, z) = \beta \frac{1}{p + \beta \gamma} \left( c_n^* - \left( \sum_{i=0}^{n-1} a_i a_{n-1-i} \right)^* (p, z) \right), \quad (24)$$

#### IV. SOLUTIONS FOR ZERO AND N-TH APPROXIMATIONS

The distributions of adsorption concentration in gas phase  $c_0(t, z)$ , the temperature of the layer  $T_0(t, z)$  and the concentration of the adsorbate (absorbed substance) in the nanopores of the adsorbent  $a_0(t, z)$  are looks like:

$$\begin{aligned} c_0(t, z) &= c_{in}(0) e^{\frac{u}{2D_{inter}} z} \Phi_c^0(t, z) + e^{\frac{u}{2D_{inter}} z} \int_0^t \frac{d}{d\tau} c_{in}(\tau) \Phi_c^0(t - \tau, z) d\tau \\ &+ c_0^0 \frac{\gamma}{1 + \gamma} \left( 1 + \frac{1}{\gamma} e^{-\beta(\gamma+1)t} - \frac{\gamma+1}{\gamma} e^{\frac{u}{2D_{inter}} z} \Phi_c^0(t, z) \right) \\ &+ \beta c_0^0 e^{\frac{u}{2D_{inter}} z} \int_0^t e^{-\beta(\gamma+1)(t-s)} \Phi_c^0(s, z) ds \end{aligned} \quad (25)$$

$$\begin{aligned} T_0(t, z) &= T_{in}(0) \Phi_T^0(t, z) + \int_0^t \frac{d}{d\tau} T_{in}(\tau) \Phi_T^0(t - \tau, z) + \\ &+ \frac{1}{\Lambda} \int_0^t \int_0^\infty \left[ HT_0^0 \mathcal{H}_T(t - \tau; z, \xi) - \right. \\ &\left. + \frac{1}{\Lambda} \int_0^\infty \int_0^\infty \left[ Q \beta \left( \mathcal{H}_T(t - \tau; z, \xi) - \right. \right. \right. \\ &\left. \left. \left. \beta \gamma \int_0^{t-\tau} e^{-\beta \gamma (t-\tau-s)} \mathcal{H}_T(\tau-s; z, \xi) ds \right) c_0^*(p, \xi) \right] d\xi d\tau \right] d\tau \end{aligned} \quad (26)$$

$$a_0(t, z) = \beta \int_0^t e^{-\beta \gamma (t-\tau)} c_0(\tau, z) d\tau \quad (27)$$

The solutions  $c_n(t, z)$ ,  $T_n(t, z)$ ,  $a_n(t, z)$  for problems (15)-(17) are the functions describing the temporal spatial distributions of adsorbent concentration in gas phase, temperature and adsorption concentration in micro and nanopores of the adsorbent particles [10, 11]:

$$\begin{aligned} c_n(t, z) &= \frac{\beta}{D_{inter}} \times \\ &\int_0^\infty \int_0^\infty \left[ \mathcal{H}_c(t - \tau, z, \xi) - \right. \\ &\left. \beta \gamma \int_0^{t-\tau} e^{-\beta \gamma (t-\tau-s)} \mathcal{H}_c(s, z, \xi) ds \right] \left( \sum_{i=0}^{n-1} a_i a_{n-1-i} \right) (\tau, \xi) d\xi d\tau \end{aligned} \quad (28)$$

$$\begin{aligned} T_n(t, z) &= \frac{Q \beta}{\Lambda} \int_0^t \int_0^\infty \left[ \mathcal{H}_T(t - \tau; z, \xi) - \right. \\ &\left. \beta \gamma \int_0^{t-\tau} e^{-\beta \gamma (t-\tau-s)} \mathcal{H}_T(s; z, \xi) ds \right] \left( \sum_{i=0}^{n-1} a_i(s, \xi) a_{n-1-i}(s, \xi) - c_n(\tau, \xi) \right) d\xi d\tau \end{aligned} \quad (29)$$

$$a_n(t, z) = \beta \int_0^t e^{-\beta \gamma (t-\tau)} \left( c_n(\tau, z) - \sum_{i=0}^{n-1} a_i(\tau, z) a_{n-1-i}(\tau, z) \right) d\tau \quad (30)$$

Here:

$$\Phi_c^0(t, z) = \frac{1}{\pi} \int_0^\pi e^{-\varphi_1(\nu)z} \frac{\sin(vt - z\varphi_2(\nu)^2)}{\nu} dv + e^{-\frac{u}{2D_{inter}} z}$$

$$\begin{aligned} \Phi_c(t, z) &= \\ &\frac{1}{2\pi} \int_0^\infty \frac{\varphi_1(\nu) \cos(vt - \varphi_2(\nu)z) + \phi_2(\nu) \sin(vt - \varphi_2(\nu)z)}{\left( \Gamma_1^2(\nu) + \nu^2 \Gamma_2^2(\nu) \right)^{1/2}} dv \end{aligned}$$

$$\Phi_T^0(t, z) = \frac{1}{\pi} \int_0^\pi e^{-\phi_1(\nu)z} \frac{\sin(vt - z\phi_2(\nu)^2)}{\nu} d\nu + e^{-\frac{u}{2D_{\text{inter}}}z},$$

$$\Phi_T(t, z) =$$

$$\frac{1}{2\pi} \int_0^\infty \frac{\phi_1(\nu) \cos(vt - \phi_2(\nu)z) + \phi_2(\nu) \sin(vt - \phi_2(\nu)z)}{\left(\Gamma_{T_1}^2(\nu) + \nu^2 \Gamma_{T_2}^2(\nu)\right)^{1/2}} d\nu,$$

$$\phi_{1,2}(\nu) = \left[ \frac{\left(\Gamma_1^2(\nu) + \nu^2 \Gamma_2^2(\nu)\right)^{1/2} \pm \Gamma_1^2(\nu)}{2} \right]^{1/2},$$

$$\Gamma_1(\nu) = \frac{u^2}{4D_{\text{inter}}^2} + \frac{\nu^2 \beta}{D_{\text{inter}}^2 (\nu^2 + \beta^2 \gamma^2)}; \quad \Gamma_2(\nu) = \frac{\nu^3 + \nu \beta^2 (\gamma + 1) \gamma}{D_{\text{inter}} (\nu^2 + \beta^2 \gamma^2)}$$

$$\phi_{1,2}(\nu) = \left[ \frac{\left(\Gamma_{T_1}^2(\nu) + \nu^2 \Gamma_{T_2}^2(\nu)\right)^{1/2} \pm \Gamma_{T_1}^2(\nu)}{2} \right]^{1/2},$$

$$\Gamma_{T_1}(\nu) = \frac{u^2 + 4\Lambda X^2}{4\Lambda^2}, \quad \Gamma_{T_2}(\nu) = \frac{H\nu}{\Lambda},$$

$$\mathcal{H}_r(\tau; z, \xi) = e^{-\frac{u_2(z-\xi)}{2}} (\Phi_r(\tau, |z-\xi|) - \Phi_r(\tau, z+\xi)).$$

$$\mathcal{H}_c(\tau; z, \xi) = e^{-\frac{u_1(z-\xi)}{2}} (\Phi_c(\tau, |z-\xi|) - \Phi_c(\tau, z+\xi)).$$

## V. NOMENCLATURE

$c$  - concentration of moisture in the gas phase in the column;  
 $a$  - concentration of moisture adsorbed in the solid phase;  $T$  - temperature of gas phase flow, °C;  $u$  - velocity of gas phase flow, m/s<sup>2</sup>;  $D_{\text{inter}}$  - effective longitudinal diffusion coefficient;  $\Lambda$  - coefficient of thermal diffusion along the columns;  $h_g$  - gas heat capacity;  $Q$  - heat sorption effect;  $H$  - total heat capacity of the adsorbent and gas;  $X^2 = 2\alpha_n / R$  - coefficient of heat loss through the wall of the adsorbent;  $R$  - radius of adsorbent of solid particles, m;  $\alpha_h$  - heat transfer coefficient;  $\gamma$  - Henry's constant;  $\beta$  - mass transfert coefficient;  $z$  - distance from the top of the bed for mathematical simulation, m;

## VI. CONCLUSION

In paper proposed theoretical foundations of mathematical modeling of nonisothermal adsorption and desorption in nanoporous catalysts for exhaust gas neutralization systems for the nonlinear Langmuir isotherm. Such approach in our

opinion most fully describes the mechanism of adsorption equilibrium for micro- and nanopore systems of the ZSM-5 zeolite. An effective linearization scheme for the nonlinear model is realized. High-speed analytical solutions of the system of linearized boundary-value problems of adsorption and desorption in nanoporous media was substantiated and obtained using Heaviside's operational method.

## REFERENCES

- [1] R.M. Barrer, "Diffusion and Flow in Porous Zeolite, Carbon or Ceramic Media, Characterization of Porous Solids", *Society of Chemical Industry*, London, 1979
- [2] B. Puertolas, M.V. Navarro, J.M. Lopez, R.Murillo, A.M. Mastral, T.Garcia "Modelling the heat and mass transfers of propane onto a ZSM-5 zeolite" *Separation and Purification Technology*. vol.86., pp.127-136, 2012
- [3] M.R.Petryk, A.M.Himich, M.M.Petryk, J.Fraissard "Modeling of heat-mass-transfer adsorption and desorption in nanoporous zeolit catalytic media of exhaust gases neutralization systems" *International scientific and technical journal "Problems of control and informatics"*, vol. 2, pp. 49-57, 2018
- [4] J.Kärger and D.Ruthven "Diffusion in Zeolites and Other Microporous Solids" *New York, John Wiley & Sons.*, 605p, 1992
- [5] J.Kärger, D.Ruthven, D.Theodorou "Diffusion in Nanoporous Materials", *Hoboken, John Wiley & Sons.*, 660 p., 2012
- [6] N.Y.Chen, T.F.Degnan and M.C.Smith "Molecular Transport and Reaction in Zeolites: Design and Application of Shape Selective Catalysis", *New York, Wiley-VCH*, 510 p., 1994
- [7] A.Prudnikov, Y.Brychkov, O.Marychev "Integrals and series. Additional chapters" – M:Nauka, 736p, 1973
- [8] I.Sergienko, M.Petryk, A.Khimith, D.Mykhalyk, S.Leclerc, J.Fraissard "Mathematical Modelling of Diffusion Process in Microporous Media (Numerical analysis and application)" *Kyiv: National Academy of Sciences of Ukraine*. 196 p., 2014
- [9] O. Heaviside "Electromagnetic Theory", *London, The Electrician*. 1-3. – E.C., 1893
- [10] M.A. Lavrentiev, B.V. Shabat "Methods of theory of functions of a complex variable", *M.:Nauka*, 736 p., 1973
- [11] M. Petryk, S. Leclerc, D. Canet, I.V. Sergienko, V.S. Deineka, J. Fraissard The Competitive Diffusion of Gases in a zeolite bed: NMR and Slice Procedure, Modelling anmd Identification of Parameters. *The Journal of Physical Chemistry C. ACS (USA)*. 119 (47). – P. 26519-26525, 2015

# Design of the Saturated Interval Experiment for the Task of Recurrent Laryngeal Nerve Identification

Mykola Dyvak, Iryna Oliynyk, Volodymyr Manzhula

Department of Computer Science, Ternopil National Economic University, UKRAINE, Ternopil, 8 Chekhova str.,  
email: mdy@tneu.edu.ua, ois@tneu.edu.ua, v.manzhula@tneu.edu.ua

**Abstract:** the method of saturated design of experiment with interval data and its application for recurrent laryngeal nerve (RLN) location identification considered in this paper. Built saturated plan of experiment makes it possible to reduce the duration of surgery by reducing the number of points of irritation woven surgical wounds to identify locations RLN and reduce the risk of damage.

**Keywords:** neck surgery, recurrent laryngeal nerve, method of design if the saturated experiment, interval analysis, interval model.

## I. INTRODUCTION

RLN monitoring by using a special neuro monitors is very important procedure during the neck surgery. These monitors work based on the principle of surgical wound tissues stimulation and estimation of results of such stimulation [1-4]. However, the monitoring procedure does not guarantee a reduction in the risk of RLN damage, but only establishes the fact of its damage (not damage). In this case, the procedures for RLN identification are actual. In the paper [5] the task of visualizing the RLN location based on evaluation the maximal amplitude of signal as response to its stimulation by alternating current was considered. In paper [6] the method of constructing the difference schemes as a model for RLN location was considered. It should be noted that the informative parameter in both methods used maximum amplitude of the signal as response to stimulation of tissues in surgical wounds. The basis for localization the RLN damaging area assigned an interval model of distribution on surgical wound surface the maximum amplitudes of information signals. However, both methods require creation the uniform grid on surgical wound for tissues stimulation, which substantially increases the time of surgical operation. Note that in both cases the amount of tissue stimulation of the surgical wound is equal to the number of nodes in the grid. Such an approach greatly increases the time of the transaction by means of the procedure for RLN identification.

Based on the case of build a mathematical model, described in [5], in this article proposed to use an method of design experiment with a minimum number of experimental points (stimulation points) in order to reduce the accuracy of localization and at the same time ensure the highest possible accuracy of RLN localization. Such plans are called saturated I<sub>G</sub>-optimal.

## II. TASK STATEMENT

The stimulation of surgical wound tissues during the neck surgery based on electrophysiological method allows to

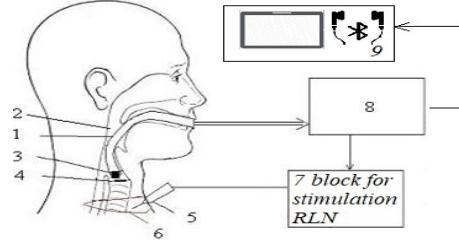
identify the type of tissue with the purpose of RLN identification.

In articles [5-6] described the method for RLN identification among tissues of a surgical wound. The scheme of this method on the fig 1 are shown.

In respiratory tube 1 that inserted into larynx 2, the sound sensor 3 implemented and positioned above vocal cords 4.

Probe 5 is connected to stimulation block. It the functions as a current generator controlled by the single-board computer 8. Surgical wound tissues are stimulated by the block 7 via probe. As a result, vocal cords 4 are stretched.

Flow of air that passes through patient's larynx, is modulated by stretched vocal cords. The result is registered by voice sensor 3. Obtained signal is amplified and is processed by single-board computer 8.



1 is respiratory tube, 2 is larynx, 3 is sound sensor, 4 are vocal cords, 5 is probe, 6 is surgical wound, 7 is block for RLN stimulation, 8 is single-board computer; 9 is output part

Fig. 1. Method of RLN identification among tissues in surgical wound

The essence of processing the signal for a given simulation point is to determine its maximum amplitude and then to construct it the interval model of distribution on surgical wound surface the maximum amplitudes. The built model makes it possible to determine the risk-sensitive area in which the RLN is localized. To construct this model, it is necessary to minimize the number of stimulations of the tissues of the surgical wound. In the research [5-6] also found that the amplitude of the information signal depends on the distance from the stimulation point to the RLN. The near this point to the RLN, the greater the value of the amplitude of the information signal.

Before the surgery on the neck, the surgeon determines the area of the surgical intervention. Assume that in our case, this area is square, given by coordinates:

$$[x_1^- \leq x_1 \leq x_1^+, x_2^- \leq x_2 \leq x_2^+] \quad (1)$$

where  $(x_1^-, x_2^-)$  - coordinates of lower limit rectangle surgery;  $(x_1^+, x_2^+)$  - coordinates of upper limit rectangle surgery.

Based on the results of works [5,6], we will assume that the values of the maximum amplitude of the information

signal - the response to the irritation of the tissues of the surgical wound will be obtained in an interval form with a constant error  $\Delta_i = \Delta, \forall i = 1, \dots, I$  for all stimulation points:

$$X = \begin{pmatrix} x_{11}, x_{12} \\ x_{21}, x_{22} \end{pmatrix}; \vec{Y} = \begin{pmatrix} [y_i^-; y_i^+] \\ \vdots \\ [y_N^-; y_N^+] \end{pmatrix}, \quad (2)$$

where  $(x_{1i}, x_{2i})$  - coordinates of stimulation points;  $[y_i^-; y_i^+]$  - interval value of maximum amplitude of information signal for  $i$ -th stimulation point/

Lets consider the mathematical model for RLN identification in kind of algebraic equation described in [5]:

$$y_o = \beta_1 \cdot \varphi_1(\vec{x}) + \dots + \beta_m \cdot \varphi_m(\vec{x}), \quad (3)$$

where  $\vec{\beta} = (\beta_1, \dots, \beta_m)^T$  is the vector of unknown parameters;  $\vec{\varphi}^T(\vec{x}) = (\varphi_1(\vec{x}), \dots, \varphi_m(\vec{x}))^T$  is the vector of known basic functions;  $\vec{x} = (x_1, x_2)$  is vector of stimulation point coordinates;  $y_o$  is predicted value of maximal amplitude of information signal in point with coordinates  $(x_1, x_2)$ .

If for any  $i$ -th stimulation point the true unknown value of information signal amplitude is  $y_{oi} = \vec{\varphi}^T(\vec{x}_i) \cdot \vec{\beta}$  belongs to the interval  $[y_i^-, y_i^+]$ , that is, we have the following condition:

$$y_i^- \leq y_{oi} \leq y_i^+, i=1, \dots, N \quad (4)$$

We get conditions:

$$\begin{cases} y_i^- \leq b_1 \varphi_1(\vec{x}_i) + \dots + b_m \varphi_m(\vec{x}_i) \leq y_i^+ \\ \vdots \\ y_N^- \leq b_1 \varphi_1(\vec{x}_N) + \dots + b_m \varphi_m(\vec{x}_N) \leq y_N^+ \end{cases} \quad (5)$$

System (5) is interval system of linear algebraic equations (ISLAE). The solution of this system will be obtained in the following form:

$$\vec{Y}^- \leq F \cdot \vec{b} \leq \vec{Y}^+, \quad (6)$$

where

$$F = \begin{pmatrix} \varphi_1(\vec{x}_1) & \dots & \varphi_m(\vec{x}_1) \\ \vdots & & \vdots \\ \varphi_1(\vec{x}_N) & \dots & \varphi_m(\vec{x}_N) \end{pmatrix}, \quad (7)$$

$\vec{Y}^- = \{y_i^-, i=1, \dots, N\}$ ,  $\vec{Y}^+ = \{y_i^+, i=1, \dots, N\}$  - vectors composed of upper and lower bounds of intervals  $[y_i^-, y_i^+]$ , respectively;  $F$  - known matrix of values for basic functions, set by expression (7). If system (6) has solutions (or one solution), then the area of these solutions is denoted by  $\Omega$ :

$$\Omega = \{\vec{b} \in R^m \mid \vec{Y}^- \leq F \cdot \vec{b} \leq \vec{Y}^+\}. \quad (8)$$

The set of all solutions  $\Omega$  of ISLAE (6) makes it possible to determine the set of equivalent (in terms of the existing interval uncertainty) interval models of static systems belonging to the functional corridor:

$$[\hat{y}(x)] = [\hat{y}^-(x); \hat{y}^+(x)], \quad (9)$$

where

$$\hat{y}^-(\vec{x}) = \min_{\vec{b} \in \Omega} (\vec{\varphi}^T(\vec{x}) \cdot \vec{b}), \quad (10)$$

and

$$\hat{y}^+(\vec{x}) = \max_{\vec{b} \in \Omega} (\vec{\varphi}^T(\vec{x}) \cdot \vec{b}), \quad (11)$$

lower and upper bounds of functional corridor, respectively.

The error of prediction of the interval model of the distribution of the maximum amplitude of the information signal will be evaluated in the following form [7-8]:

$$\Delta_{y(\vec{x})} = \max_{\vec{b} \in \Omega} (\vec{\varphi}^T(\vec{x}) \cdot \vec{b}) - \min_{\vec{b} \in \Omega} (\vec{\varphi}^T(\vec{x}) \cdot \vec{b}) \quad (12)$$

As you can see, the minimum number of stimulation points should be equal to the number of unknown  $m$  model parameters. Such experiments are called "saturated", the matrix  $X$  of the set (2) is a matrix of the plan, and the matrix  $F$  is an information matrix [7-8].

Thus, the purpose of this work is to find such a square matrix of a plan or an informational square matrix  $F_m$  of a saturated experiment that, at a known interval error  $\Delta_i = \Delta, \forall i = 1, \dots, m$ , would provide the best of the predictive properties of the interval model of the distribution of the maximum amplitude of the information signal. This condition will ensure the highest precision localization RLN using this model.

### III. METHOD OF DESIGN I<sub>G</sub>-OPTIMAL SATURATED EXPERIMENT

In the case of using a "saturated" block to estimate the parameters of interval model, its minimum prediction error in the input variable area  $\vec{x} \in \chi$  is reached at one of the points of a given set of input variables  $\vec{x}_j$  ( $j=1, \dots, m$ ) [7-11]:

$$\begin{aligned} \Delta_{\min} &= \min_{\vec{x}_j, j=1, \dots, m} \{ \hat{y}^+(\vec{x}_j) - \hat{y}^-(\vec{x}_j) \} = \\ &= \min_{j=1, \dots, m} \{ 2\Delta_j \} \end{aligned} \quad (13)$$

$$\vec{x}^{\min} = \arg \min_{x_j, j=1, \dots, m} \{ \hat{y}^+(\vec{x}_j) - \hat{y}^-(\vec{x}_j) \} \quad (14)$$

Procedures for calculating the maximum prediction error by interval models are much more complicated, even in the case of estimating its parameters based on the "saturated" block of ISLAE [7-11].

In work [8] expressions are given to calculate the error value at any point for the specified event:

$$\Delta_{y(\vec{x})} = 2 \cdot \sum_{j=1}^m |\alpha_j(\vec{x}) \cdot \Delta_j|, \vec{x} \in \chi \quad (15)$$

$$\bar{\alpha}^T(\vec{x}) = \vec{\varphi}^T(\vec{x}) \cdot F_m^{-1} \quad (16)$$

where  $\alpha_j(\vec{x})$  -  $j$ -th component of vector  $\bar{\alpha}(\vec{x})$ , which in the general case depends on the choice of a point on the experiment area;  $\Delta_j = 0,5 \cdot (y_j^+ - y_j^-)$  - interval errors in  $\vec{x}_j$  observation points.

Based on the expressions (15), (16), we formulate a condition for choosing a "saturated" block to minimize the maximum prediction error in the area of the values of the input variables that determine the matrix  $X$  rows:

$$\max_{\vec{x} \in X} \left\{ 2 \cdot \sum_{j=1}^m |\alpha_j(\vec{x}) \cdot \Delta_j| \right\} \xrightarrow{F_m} \min, \quad (17)$$

$$\vec{\alpha}^T(\vec{x}) = \vec{\varphi}^T(\vec{x}) \cdot F_m^{-1}$$

As you can see, the maximum value of prediction error in the area of input variables  $\vec{x} \in X$  depends on the chosen "saturated" block. We will denote it:  $\Delta_{\max}(F_m)$ .

Expression (17) provides minimization of the maximum error of prediction of the interval model among all possible "saturated" blocks selected from ISLAE (5).

Let's make an analogy with the theory of design successive  $I_G$ -optimal interval experiment plans that minimize the maximum error of prediction of interval models [7-11]. In our case, the essence is: design some series of experiments with a small amount of observations (for example, a saturated experiment); get the corridor of interval models; analysis of the predictive properties of these models and on this basis the design of the next one observation [7-11].

Considering the requirements of providing optimal prognostic properties of the interval model (minimizing the maximum of prediction error) in the area of input variables, it is advisable to use this approach to select the "saturated" block from ISLAE (5) in order to simplify the task (17).

Note that in the procedure of  $I_G$ -optimal design on the first iteration, the "saturated" block is also chosen according to the  $I_G$ -criterion, the expression for which is represented (17). In our case, such iteration is meaningless due to high computational complexity. Therefore, in the first step of the method of estimating the set of values of the parameters of the interval models of static systems, the "saturated" block of ISLAE will be chosen arbitrarily.

Let the structure of the mathematical model of the static system be defined by the expression (3) with unknown parameters, given interval data (4) and ISLAE formed in the form (5).

Choose from ISLAE (5) an arbitrarily "saturated" block and compute its area of solutions, construct a prediction corridor with interval models in the form (9), where  $\Delta_{\tilde{y}(\vec{x})}$  are defined by expressions (15), (16).

By analogy with the procedure of successive  $I_G$ -optimal design, based on the expressions (15), (16), among ( $i=1, \dots, N$ ) rows of the matrix  $X$  values of the input variables for which ISLAE (5) is composed, choose the vector-row  $\vec{x}^{\max}$  for which we calculate the maximum prediction error, that is:

$$\vec{x}^{\max} = \arg \max_{\vec{x}_i, i=1, \dots, N} \left\{ 2 \cdot \sum_{j=1}^m |\alpha_j(\vec{x}_i) \cdot \Delta_j|, \vec{x}_i, i = 1, \dots, N \right\}, \quad (18)$$

$$\vec{\alpha}^T(\vec{x}_i) = \vec{\varphi}^T(\vec{x}_i) \cdot F_m^{-1}$$

The vector obtained from expression (18) is a vector of values of input variables, which defines a certain interval equation in ISLAE (5). According to the procedure of successive  $I_G$ -optimal design, it is necessary to carry out the following measurement with respect to this vector-row.

Based on expression (14) to determine the vector of values of input variables, where the prediction error is minimal in the experiment area, we can state that if the vector  $\vec{x}^{\max}$  coincides with the vector of values of input variables, for which one of the interval equations of the "saturated" block

of ISLAE is constructed, then it would set the point with the minimum value of the prediction error. It is advisable to replace one of the interval equations in the ISLAE(5) by interval equation with the vector of values of the input variables  $\vec{x}^{\max}$  defined by the expression (18) in the current "saturated" block. By analogy with the procedure of successive  $I_G$ -optimal design, we simulate the additional measurement procedure for the vector-row  $\vec{x}^{\max}$  with the maximum error of prediction of the interval model, obtaining measurements with a minimum interval error according to the expression:

$$\max_{\vec{x} \in X} \left( \Delta \cdot \sqrt{\vec{\varphi}^T(\vec{x}) \cdot (F_m^T \cdot F_m)^{-1} \cdot \vec{\varphi}(\vec{x}) \cdot m} \right) \xrightarrow{F_m} \min. \quad (19)$$

However, in contrast to the  $I_G$ -optimal sequential design procedure of an experiment, we choose a given point on a discrete set of vector-rows  $\vec{x}_i$  ( $i = 1, \dots, N$ ) of matrix  $X$ . Denote the lower and upper bound for the resulting interval by  $[\hat{y}_{\min}^-; \hat{y}_{\min}^+]$ .

We will carry out the above procedure for each interval equation of the "saturated" block. We get  $p$  new "saturated" blocks ( $p=m$ ). As a result, for each of the  $m$  "saturated" blocks we obtain  $m$  values of maximum errors for the corresponding interval models:

$$\Delta_{\max}(F_m(p)) = \max_{x_i, i=1, \dots, N} \left\{ 2 \cdot \sum_{j=1}^m |\alpha_{jp}(\vec{x}_i) \cdot \Delta_j| \right\}, \quad (20)$$

$$\vec{\alpha}_p^T(\vec{x}_i) = \vec{\varphi}^T(x_i) \cdot F_m^{-1}(p), p = 1, \dots, m$$

where  $p$  is index, which means a number of "saturated" block,  $F_m(p)$  is matrix of basic functions values for  $p$ -th block,  $\alpha_{jp}(\vec{x}_i)$  is  $i$ -th component of vector  $\vec{\alpha}$  for  $p$ -th "saturated" block.

To choose the optimal "saturated" block in this step, instead of a complex computational procedure (17), it is sufficient to choose from the  $m$  "saturated" blocks the one that provides the lowest value of the sequence (23), that is:

$$F_m^{opt} = \arg \min_{p=1, \dots, m} \{ \Delta_{\max}(F_m(p)), p = 1, \dots, m \}, \quad (21)$$

Applying procedure (18), we obtain  $\vec{x}^{\max}$  - the point at which the maximum error of the prediction by the interval model, the estimation of the set of values of parameters which is calculated from the chosen "saturated" block in the above-described method. Further iterations are continued until such a "saturated" block is obtained, the replacement of interval equations which does not lead to a decrease in the maximum prediction error by interval models.

The algorithm of realization of proposed method described below [8]:

Step 1. Choose the "saturated" block of ISLAE (5).

Step 2. Determination of the vector-row  $\vec{x}^{\max}$  of the matrix  $X$  for solving the task (18).

Step 3. The iteration of the phased replacement of each of the  $m$  interval equations of the "saturated" block on the ISLAE (5) to the interval equation with a vector of values of input variables  $\vec{x}^{\max}$  (forming a set of "saturated" blocks) and calculating maximum prediction errors (15) for interval

models.

Step 4. Choose the optimal "saturated" block based on expression (21).

Transitioning to step 2.

Note that when transitioning to step 2, for a received "saturated" block, the estimations of the set of parameters of the interval model and the maximum error of the interval model constructed for this block will be known.

We implement a sequence of steps until the "saturated" block is received at the last step, any replacement of interval equations does not lead to a decrease in the maximum prediction error for the interval-based models constructed on its basis.

#### IV. EXAMPLE OF APPLICATION THE PROPOSED METHOD FOR RLN IDENTIFICATION

The example of constructing a model of distribution on the surface of a surgical wound the maximum signal amplitudes as reaction to stimulation of surgical wound tissues described in article [5]. The structure of this mathematical model, obtained from the work [5], has the following form:

$$y(\vec{x}) = b_0 + b_1 \cdot \sin^2(x_1 \cdot x_2 \cdot \frac{\pi}{36}) + b_2 \cdot x_2 + b_3 \cdot (x_2^2) \quad (24)$$

A fragment of data obtained during the surgical operation on thyroid gland is given in Table 1 in [5].

We apply the method of selection the "saturated" block from ISLAE to find the most informative points.

As a result, the coordinates of the points of the saturated plan are found: [1.5;3], [3;6], [6;0.5], [6;6] (Fig. 2).

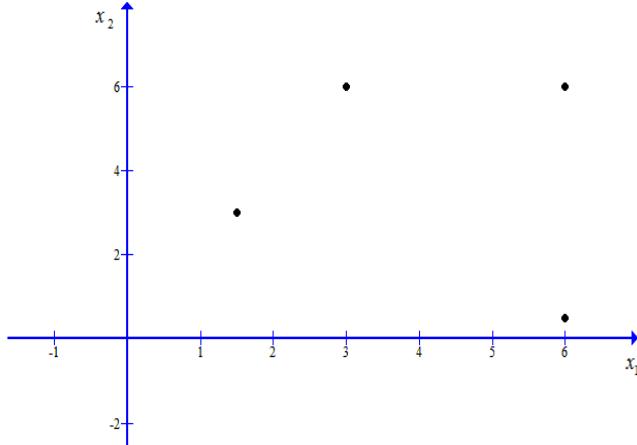


Fig. 2. The coordinates of the points found in the saturated plan for the RLN identification task

#### V. ACKNOWLEDGMENT

This research was supported by National Grant of Ministry of Education and Science of Ukraine "Mathematical tools and software for classification of tissues in surgical wound during surgery on the neck organs" (0117U000410).

#### VI. CONCLUSIONS

The method of design the saturated experiment with interval data on the principles of formation of a saturated block of the interval system of linear algebraic equations, as

well as its application for the problem of locating RLN in the process of operation on the neck organs considered in this article. This saturated experimental plan allows to reduce the duration of surgical surgery by reducing the number of imitation points of a wound surgical wound to detect a RLN. If for the cases considered in other works construct a grid of  $m^2$  stimulation points, then in the case of a saturated plan, the number of points is equal to the number of unknown coefficients of the model, which is tens of times smaller.

#### REFERENCES

- [1] M.C.D. Poveda, G. Dionigi, A. Sitges-Serra, M. Barczynski, P. Angelos, H. Dralle, E. Phelan and G. Randolph. "Intraoperative Monitoring of the Recurrent Laryngeal Nerve during Thyroidectomy: A Standardized Approach (Part 2)". *World Journal of Endocrine Surgery*, vol. 4, no. 1, pp. 33-40, 2012.
- [2] V.K. Dhillon and R.P. Tufano. "The pros and cons to real-time nerve monitoring during recurrent laryngeal nerve dissection: an analysis of the data from a series of thyroidectomy patients". *Gland Surgery*, vol. 6, no. 6, pp. 608-610, 2017.
- [3] H.Y. Kim, X. Liu, C.W. Wu, Y.J. Chai and G. Dionigi. "Future Directions of Neural Monitoring in Thyroid Surgery". *Journal of Endocrine Surgery*, vol. 17, no. 3, pp. 96-103, 2017.
- [4] W.E. Davis, J.L. Rea, J. Templer., "Recurrent laryngeal nerve localization using a microlaryngeal electrode". *Otolaryngology – Head and Neck Surgery*, vol. 87, no. 3, pp. 330-333, 1979.
- [5] M. Dyvak, O. Kozak, A. Pukas. "Interval model for identification of laryngeal nerves". *Przeglad Elektrotechniczny*, vol. 86, no. 1, pp. 139-140, 2010.
- [6] N. Porplysya, M. Dyvak. "Interval difference operator for the task of identification recurrent laryngeal nerve", *16th International Conference On Computational Problems of Electrical Engineering (CPEE)*, pp. 156-158, 2015.
- [7] M. Dyvak, I. Oliynyk. "Estimation Method for a Set of Solutions to Interval System of Linear Algebraic Equations with Optimized "Saturated Block" Selection Procedure". *Computational Problems of Electrical Engineering*, Lviv, 2017. – V. 7, No. 1. – P. 17-24.
- [8] Dyvak M., "Tasks of mathematical modeling the static systems with interval data", *Economic thought*, Ternopil, 2011, 216 p. (in Ukrainian).
- [9] Wu C. F. J., Hamada M. S., "Experiments: Planning, Analysis and Optimization", Wiley, 2009, 743 p.
- [10] Shary S.P., "Algebraic Approach to the Interval Linear Static Identification, Tolerance, and Control Problems, or One More Application of Kaucher Arithmetic", *Reliable Computing* 2(1) (1996), p. 3–33.
- [11] Walter E. and Pronzato L., "Identification of parametric model from experimental data". London, Berlin, Heidelberg, New York, Paris, Tokyo: Springer, 1997. – 413 p.

# Discrete Dynamic Model of Retail Trade Market of Computer Equipment in Ukraine

Mykola Dyvak<sup>1</sup>, Vasyl Brych<sup>2</sup>, Iryna Spivak<sup>3</sup>, Lyudmyla Honchar<sup>4</sup>, Nataliya Melnyk<sup>5</sup>

1. Department of Computer Science, Ternopil National Economic University, UKRAINE, Ternopil, 8 Chekhova str., email: mdy@tneu.edu.ua

2. Department of International Tourism and Hotel Business, Ternopil National Economic University, UKRAINE, Ternopil, 11 Lvivska str., email: v.brych@tneu.edu.ua

3. Department of Computer Science, Ternopil National Economic University, UKRAINE, Ternopil, 8 Chekhova str., email: i.spivak@tneu.edu.ua

4. Department of Computer Science, Ternopil National Economic University, UKRAINE, Ternopil, 8 Chekhova str., email: [l.honchar@tneu.edu.ua](mailto:l.honchar@tneu.edu.ua)

5. Faculty of Economics, Ivan Franko National University of Lviv,  
UKRAINE, Lviv, email: talya\_m@ukr.net

**Abstract:** In this article the discrete dynamic model of the retail computer market functioning in Ukraine is considered. The existing distribution models are analyzed between the main entities in the market. A new model is proposed and an example of the IT market dynamics is shown.

**Keywords:** retail computer market; discrete dynamic model; correction function; predicted indicators.

## I. INTRODUCTION

Mathematical modeling is one of the most important tools for researching economic processes. The constant demand for new IT generates new types of computer equipment (CE), as well as new IT services. These proposals are aimed not only at meeting the needs of business structures, but also at an average household information consumer. The last circumstance stimulates the rapid development of CE retail market, on which the main buyer is the individual consumer of the information product. To predict the development of this market, it is necessary to build its dynamics model. Such models are described in the works [1, 2], where the problems of their structural and parametric identification on the basis of data analysis are considered. Such models are called discrete dynamic or differential operators [1].

To construct a model of market dynamics, let's consider the important moments of the subject modeling area. At the domestic CE retail market we distinguish four sellers categories: consumer electronics, specialized computer stores, mobile communication stores and B2B-sector enterprises [3]. A narrow range of sellers in the domestic market of CE is conditioned by the monopoly in this area. Those structures, which operate on the domestic market, play the role of a distributive link, which does not define strategic directions of IT development and applies to advance achievements in this field. Without having their own product and struggling for a part of the market share they are forced to behave extremely responsibly in their own business, paying attention to a number of factors, on which they have no influence [3].

The need of planning strategies and tactics of doing business, challenges the participants with complex problems of mathematical modeling of processes in the retail market of

CE, which are happening in this market. The modeling results can be taken into consideration during development and implementing business policies in specific circumstances. Predicted by discrete dynamic mathematical models, the indicators of business development make it possible to adequately assess their own investment opportunities and to attract investments from the side. Mathematical models are used to track trends in the market. This allows you to correctly emphasize the position of advertising companies in order to timely implement effective marketing activities. Consequently, the skillful use of the market's subject methods of mathematical modeling in the final result gives you a number of advantages in the competition [3].

Rapid changes in the IT-industry trends put high demands to the possibility of mathematical models. For example, a model that adequately reproduces sales of storage devices, will not necessarily work in the case of a rapid and massive transition from the use of optical disks to electronic media. This can be explained by the fact that entities in the common market react differently to these changes. Some of them, which are oriented to the sale of goods in large batches, are not able to quickly abandon the devices, which action is based on "outdated" technologies, since there are a large number of such devices in the warehouses. Obviously, the advertising and marketing policies of such structures will be aimed at reducing their stocks as quickly as possible. Other market players, who are more mobile in the process of transitioning to new technologies, are pursuing a policy aimed at promoting the latest devices. Therefore they are receiving competitive advantages that are not taken into account in conventional foreseen models [6].

It is also possible that new sellers will enter the market, who bet on the latest IT technology, or some vendors will be replaced by others. Individual sellers can change their priorities and refuse to commerce certain types of CE. Apparently, such vendor substitutions also require correction of existing linear dynamic models [3], for example, by introducing a nonlinear part that reflects switching processes.

The purpose of this study is to develop a new model of the retail market of CE based on the rapid changes that take place

in IT. To achieve the goal, you need to solve these problems: analysis of existing models of distribution of the retail market of CE between sellers by major segments; to construct a discrete dynamic model with a "switch" for describing the CE market, which is subjected to change of circumstances of its functioning; check the model for adequacy.

## II. TASK DEFINING

Work [3] proposes a model that reflects the distribution of the retail CE market dynamics between the four large suppliers (entities) by separate segments of the market. Each segment corresponds to a certain type of CE. In the example given in [3], four segments are considered, namely: personal computer segment (PC), laptops segment, a segment of displays and a segment of multifunctional devices (MFD).

To simulate the dynamics, the mathematical model is selected as a linear discrete equation (differential operator) in the form [3,6]:

$$\begin{cases} \vec{x}^{(k+1)} = F \cdot \vec{x}^{(k)} + G \cdot \vec{v}^{(k)} \\ \vec{y}^{(k+1)} = C \cdot \vec{x}^{(k+1)}, k = 0, 1, 2, \dots \end{cases} \quad (1)$$

where  $\vec{x}^{(k)}$  – vector of state variables, which characterize the change in the formal state of the market;  $\vec{v}^{(k)}$  – vector of input variables, which reflect the effect of factors on the market;  $\vec{y}^{(k)}$  – vector of output variables, which reflect the characteristics of the distribution of the market among its subjects;  $k$  – time sequence number, in which the value of the components of the corresponding vectors is determined;  $F, G, C$  – valid matrices of the corresponding measurements.

Parameters of this model (matrix elements  $F, G, C$ ) are obtained in the form of parametric identification according to the well-known Ho-Calman algorithm [4-6]. The reason of parametric identification is the Henkel block matrix. Each block of this matrix in the case [3] is formed from the data on the market share, which is occupied by the  $j$ -th element ( $j = \overline{1, 4}$ ) in the  $i$ -th ( $i = \overline{1, 4}$ ) market segment by the results of  $k$ -th year ( $k = \overline{1, 4}$ ).

In the work [3] an assumption was made, that conditions of functioning of the retail market CE were unchanged. Model built in this way, allowed to get a rough estimate of the distribution of the market for  $k+1$  period among the main categories of vendors for each segment of the market.

## III. IMPROVED DISCRETE DYNAMIC MODEL

As shown in [3], even with the preservation of previous trends in the functioning of the retail market of CE, the full adequacy of the basic model could not be achieved. Only use of optimization procedures allowed to get adequate values for predicted indicators. Obviously, when market trends fluctuate, the proposed model becomes inadequate. Therefore, it is necessary to complicate the structure of the model, taking into account changes in the market. Such a complication can be done using the methods of structural identification [2]. However, in our case, the dynamics model

is deterministic and its structure can be obtained based on a detailed analysis of the subject area. In order to take into account the changes in trends changes is offered to modify the proposed model by introducing switching functions that simulate a sharp change in market conditions.

Let's consider the case when at a certain point in time  $t_{out}$  on the market in a particular segment changes the number of subjects, namely, one of the subjects leaves the specified segment. This means that the number of non-zero vector components  $\vec{y}^{(k)}$  in (1) decreases. To represent this possible change we introduce into the equations of the output variables of model (1) an additional diagonal matrix  $T$ , which has such a general view:

$$T = \begin{pmatrix} f(t_1) & 0 & \cdots & 0 \\ 0 & f(t_2) & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & f(t_n) \end{pmatrix}, \quad (2)$$

where is the switch function  $f(t_s) = \begin{cases} 1, & t_s < t_{out}, \\ 0, & t_s \geq t_{out} \end{cases}, s = \overline{1, n}, n -$  number of vector components  $\vec{y}^{(k)}$  in the model (1). Then we will get:

$$\begin{cases} \vec{x}^{(k+1)} = F \cdot \vec{x}^{(k)} + G \cdot \vec{v}^{(k)} \\ \vec{y}^{(k+1)} = T \cdot C \cdot \vec{x}^{(k+1)}, k = 0, 1, 2, \dots \end{cases} \quad (3)$$

Based on these tasks in [3], we will analyze the obtained model structure. At the same time, we assume that sellers belonging to a certain category, namely: mobile communication stores, refused to sell monitors. In fact, this situation was observed in 2015. In model (3) this year corresponds to the order number of the time point  $k = 5$ . And the serial number of the monitor –  $s = 3$ . So with  $k = 5$  we get  $t_3 = t_{out}$  and accordingly the switching function  $f(t_3) = 0$ , and the corresponding matrix  $T$  will have the form:

$$T = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \quad (4)$$

According to model (3), foreseen indicators reflecting the the distribution of CE market in the segment of monitors between entities, which are left in this segment are: consumer

electronics – 38,5%; specialized stores – 23,7%; B2B sector – 36,2%.

Obviously, that the market share, which according to forecasts should have mobile exhibition halls (which is 1,6%), were distributed among the remaining enterprises in the segment of monitors. Assuming that this division took place in proportion to that part, that each subject has in this segment, then each of them received an additional share in the amount respectively 0,62%, 0,38% i 0,6%. Consequently, the final forecast for the distribution of the retail market of monitors is as follows: consumer electronics – 39,12%; specialized stores – 24,08%; B2B - sector – 36,8%.

If we apply the optimization procedure for the initial distribution (38,5%; 23,7%; 36,2%) and under conditions, that the goal function as the sum of the shares of all sellers remaining in the segment is equal 100%, and the deviation of each particle does not exceed 1,6%, then we obtain the following final values: consumer electronics – 36,9%; specialized stores – 25,3%; B2B - sector – 37,8%.

The difference between the two methods of estimating the forecasted values of market distribution lies within the limits [2,22%; -1%]. Obviously, taking into account such clear limitations in forecasting, the entity of the retail market CE has the opportunity to more precisely define its own business development strategy.

To reproduce changes in the retail market of CE, which are connected with elimination or gradual abandonment of this market by separate subjects, into the model (1) it is appropriate to enter correction functions into the vector of input variables  $\vec{v}^{(k)}$ . Let's look at an example where subject 3 (mobile stores) dramatically reduces its presence on the market in all its segments. Instead, its market share is individually trying to be captured the subject 2 (specialized stores).

According to the Ho-Calman algorithm [4-6] model (3) reproduces indicators of a particular subject, on the basis of which it was built, on the condition, that at the starting moment of time ( $k=1$ ) the corresponding input variable is equal to 1, and all others - equal to zero. So for the second and third subject, the vectors of the input variables are respectively:

$$\vec{v}^{(1)} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \text{ i } \vec{v}^{(1)} = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}. \quad (5)$$

For all other moments of time ( $k = 2,3,\dots$ ) all components of the input vector must be zero.

Reduction of the share in all segments in the market of an individual entity simulating by replacing the value of the corresponding component in the vector  $\vec{v}^{(1)}$ , namely: the value 1 is replaced by the value  $\alpha$  ( $0 < \alpha < 1$ ). The smaller the value  $\alpha$ , the faster is the process of leaving this subject of the CE market.

Vice versa, capturing part of the share market by specific subject, which previously belonged to another entity, are described by replacing the value of the corresponding component in the vector  $\vec{v}^{(1)}$  on  $\beta$  ( $\beta > 1$ ).

For a general case let's write a vector  $\vec{v}^{(1)}$  with the correction function  $g(\xi_1, \xi_0)$ , where  $\xi_0$  i  $\xi_1$  – respectively market shares, which the subject takes before and after changing certain condition. Finally vector  $\vec{v}^{(1)}$  will have the following generalized form:

$$\vec{v}^{(1)} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ g(\xi_1, \xi_0) \\ \vdots \\ 0 \end{pmatrix}, \quad (6)$$

$$\text{where } g(\xi_1, \xi_0) = \begin{cases} 1, & \xi_1 = \xi_0 \\ \alpha, & \xi_1 < \xi_0 \\ \beta, & \xi_1 > \xi_0 \end{cases}.$$

Therefore, we'll assume, that subject 3 has rapidly reduced its presence in the CE retail market. Let this be decreased 70%. This means that in this case, the correction function  $g(\xi_1, \xi_0)$  has value  $\alpha = 0,3$ . By introducing such a value into the main model (1) we obtaining foreseen market shares in different segments for this subject, namely: PC segment – 0,4%, laptop segment – 3,2%, displays segment – 0,4%, MFD segment – 1,6%.

If entity 2 can individually capture that market share, which entity 3 has left, so this means, that the correction function for it considering his total share for all segments will have value  $\beta = 1,15$ . As a result of foreseeing, we obtain the following values for different segments of the market: PC segment – 9,4%, laptop segment – 24,2%, displays segment – 27,3%, MFD segment – 24,4%.

For such values  $\alpha$  and  $\beta$ , segments modeling error are within the range limits [0,6%;4,3%].

In terms of practice, it is unlikely, that entity 2 alone captures all market shares, which entity 3 has left. Most likely, this circumstance will be used by other entities. In our case, this is the subject 1 (consumer electronics) and subject 4 (the enterprise B2B sector). If we assume, that released market share by subject 3 is divided between subjects 1, 2 i 4 in proportions 5:2:3, so the corresponding values of the correction functions will be:  $\beta_1 = 1,35$ ,  $\beta_2 = 1,14$ ,  $\beta_3 = 1,21$ . For such correction function values, after optimization procedures use by the method [7-11] with restriction, which are within [0,4%;4,5%], we get a forecast of market distribution by segments.

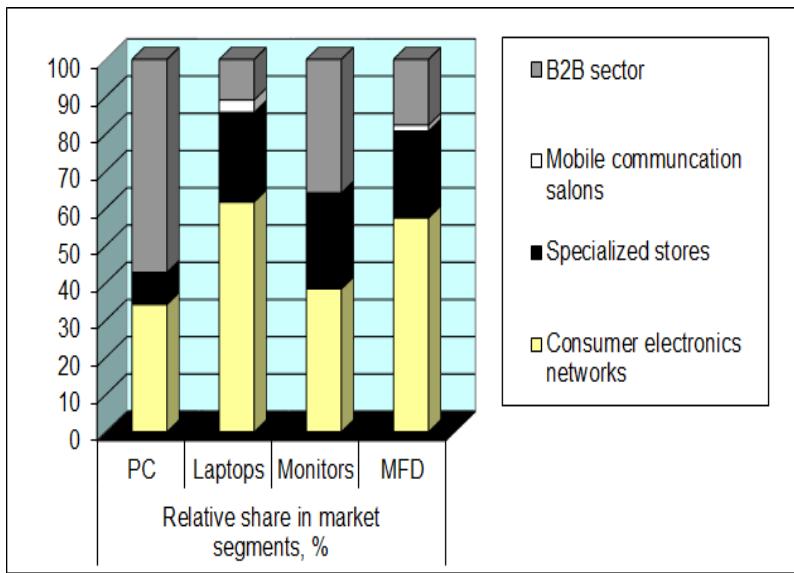


Fig. 1. Estimated distribution of the CE retail market.

Consequently, the introduction into the model (1) the correction function allows to adequately describe the processes in the retail market of CE in case of trends changes.

#### IV. CONCLUSION

The instability of the situation on the retail market of CE due to various factors, in particular the rapid development of IT. This complicates the simulation of those processes, which are happening on it, since existing models do not count technological changes in IT industry, and corresponding changes in the structure of the market. Therefore, these models should be modified by introducing correction functions to them.

The correctional functions proposed in this research, make it possible to predict the distribution of the retail CE market for  $y$  segments between entities in the case of abrupt changes in the tendencies of its functioning. Namely: in the case of a certain subject's refusal to trade in a particular type of CE, and in the case of a rapid decrease in the presence of the entity in all segments of the market.

In the future studies non-linearity of segmental redistribution of the market between entities should be taken into account. To do this, you need to switch to a more complex form of the main model, for example, bilinear. However, such approach is suitable not for all cases. Therefore, in further researches, the structure identification methods based on inductive approach will be used to choose the model [12]. Also, it is advisable to take into account the uncertainty of given data, scilicet, their variety on different intervals.

#### REFERENCES

- [1] I. Voytyuk, M. Dyvak, V. Spilchuk, "Research of quality characteristics of models structure in kind of interval difference operator", Proceedings of International Conference CADSM'2011, Polyana-Svalyava, 2011, pp. 87.
- [2] N. Porplytsya, M. Dyvak, I. Spivak, I. Voytyuk, "Mathematical and algorithmic foundations for implementation of the method for structure identification of interval difference operator based on functioning of bee colony", Proceedings of International Conference CADSM'2015, Lviv, Ukraine, 2015, pp. 196-199.
- [3] N. Melnyk, M. Dyvak, "Modeling of the dynamics distribution model by retail computer market segments", Lviv national university herald. Economical series, Lviv, 2016, Edition 53, pp. 150-157.
- [4] R.E. Kalman, P.L. Falb, M.A. Arbib, Topics in mathematical system theory. McGraw Hill Book Co, 1969.
- [5] Bowden, R.: The theory of parametric identification. Econometrica 41, 1069–1074 (1973).
- [6] Graupe, D.: Identification of systems. Technology & Engineering (1976).
- [7] P.H. Stakhiv, Y.Y. Kozak, O.P. Hoholyuk, "Discrete macromodeling in electrical engineering and related fields", monograph: Lviv Polytechnic Publishing House, 2014. – 260 p.
- [8] E. Rosolowski, P. Stakhiv, O. Hoholyuk, "Prospects of discrete macromodels usage for calculation of electric power systems modes", Modern Problems of Radio Engineering, Telecommunications and Computer Science, Proceedings of the 13th International Conference on TCSET' 2016, Lviv-Slavsko, p.55-57.
- [9] Rastrigin, L.A.: A random search. Znanie, Moscow (1979). (in Russian)
- [10] Rastrigin L.A.: Adaptation of complex systems. Zinatne, Riga (1981). (in Russian)
- [11] I. Calishchuc, M. Dyvak, P. Stakchiv, "Identyfikacja dynamicznego modelu obwodu elektrycznego na podstawie danych interwałowych", Przegląd Elektrotechniczny, Nr. 2/2005, pp. 60-62.
- [12] Ivakhnenko A.G., "The inductive method for self-organization of models of complex systems" Naukova Dumka, 1981, 296pp. (in Russian).

# Methods and Tools for Electrophysiological Monitoring of Recurrent Laryngeal Nerve Monitoring During Surgery on Neck Organs

Mykola Dyvak<sup>1</sup>, Volodymyr Tymets<sup>1</sup>, Andriy Dyvak<sup>2</sup>, Oksana Huhul<sup>1</sup>,

1. Faculty of Computer Information Technologies, Ternopil National Economic University, UKRAINE, Ternopil, 8 Chekhova str., e-mail: [mdy@tneu.edu.ua](mailto:mdy@tneu.edu.ua), [volodymyrtymets@gmail.com](mailto:volodymyrtymets@gmail.com), [oksansggg@i.ua](mailto:oksansggg@i.ua)

2. Department of Surgery with Urology №1 by L.Ya. Kovalchuk, I. Horbachevsky Ternopil State Medical University, UKRAINE, Ternopil, 1 Maidan Voli, e-mail: [dyvak\\_anmy@tdmu.edu.ua](mailto:dyvak_anmy@tdmu.edu.ua)

**Abstract – Methods and tools for electrophysiological monitoring and identification of recurrent laryngeal nerve (RLN) are considered in the paper. The method of identification and tools for stimulation of surgical wound tissues during surgery on neck organs are represented. Improved information technology of RLN identification and results of its application are shown.**

**Keywords – neck organs surgery, recurrent laryngeal nerve, single-board computer, multi-functional electro-stimulator.**

## I. INTRODUCTION

Recurrent laryngeal nerve (RLN) monitoring is very important procedure during the neck surgery. For this purpose, special neuro monitors are used. They work based on the principle of surgical wound tissues stimulation and estimation of results of such stimulation [1-5]. The main problem that arises during this process is the proper choice of stimulation methods. In [1] and [4], the latest results of researches related to RLN neuro-monitoring are represented.

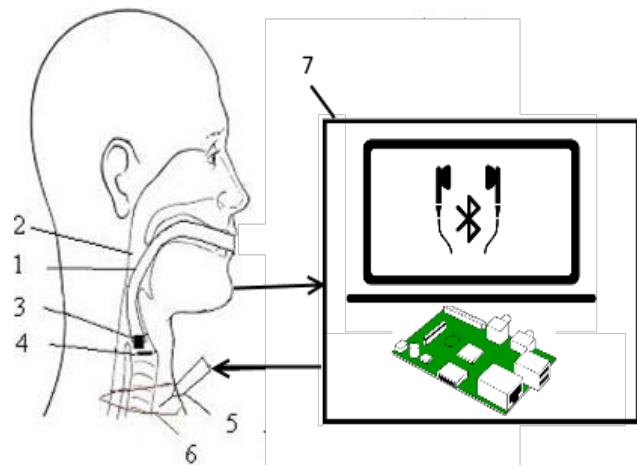
Other electrophysiological method of RLN stimulation and monitoring requires the alternating current with fixed frequency [7]. The methods and mathematical models for solving this problem are considered in [6-7].

The main problem of the mentioned methods is a high risk of RLN damage. This risk is mainly caused by the accuracy of output information signal (result of surgery wound tissues stimulation) processing. Methods of spectral analysis of this signal [7] give an opportunity to select the main spectral component and to classify the surgery wound tissues. Methods of RLN location visualisation based on building the model of distribution of amplitude of the information signal on the surgery wound allow to detect the high-risky area [7]. The combination of these methods in one information technology of processing of signal (reaction on the surgery wound tissues stimulation) is an important task. The paper is dedicated to this task.

The method of RLN monitoring is based on the task of its stimulation as the first sub-task. Other aspect of the task is the processing of reaction on RLN stimulation. After the processing, a conclusion about the RLN location in the surgery area is made.

## II. TASK STATEMENT

Let's consider the functionality principles of existing hardware solution for RLN identification. The scheme of the device is shown in Fig.1 [7].



1 is respiratory tube, 2 is larynx, 3 is sound sensor, 4 are vocal cords, 5 is probe, 6 is surgical wound, 7 is multifunctional block for RLN stimulation and visualization of RLN

Fig. 1. Scheme of device and method for RLN identification

In respiratory tube 1 that is inserted into larynx 2, the sound sensor 3 is implemented and positioned above vocal cords 4.

Probe 5 connected to alternator. It performs the function of a current generator controlled by the single-board computer 7. Surgical wound tissues are stimulated by the alternating current with the fixed frequency via probe. As a result, vocal cords 4 are stretched.

Flow of air that passes through patient's larynx, is modulated by stretched vocal cords. The result is registered by voice sensor 3. Obtained signal is amplified and is processed by single-board computer 7.

For processing of obtained signal, special software is installed on a single-board computer. The main functions of the software are:

- segmentation of information signal based on analysis of its amplitude;

- analysis of amplitude spectrum using Fourier-transform;
- calculation of a spectral component with a maximal amplitude (let's call it "the main spectral component" further);
- classification of tissues of surgical environment at the points of stimulation using threshold method.

Software for changing the frequency of RLN stimulation is written in programming language Node.js. Node was created by Ryan Dahl. Now Node.js is a trademark of Joyent, Inc. and is used with its permission and maintained by the Node.js Foundation [8-9]. Node is open-source platform and located on Github.

Block of processing and displaying information is developed based on single-board computer Raspberry Pi 3 [10].

Raspberry Pi 3 was selected because of two upgrades made to it. The first one is a next generation Quad Core Broadcom BCM2837 64-bit ARMv8 processor making the processor speed increased from 900 MHz in the Pi 2 to up to 1.2GHz in the Pi 3.

The second one is addition of the built-in BCM43143 Wi-fi chip. There is also Bluetooth Low Energy (BLE) module implemented on the board.

At the same time, such an approach does not ensure a significant decrease in the RLN damage risk. For detection of high-risky area of surgery, it is expedient to build a mathematical model for RLN identification.

For these purposes, we use the properties of surgical wound tissues which are characterized by a different reaction on the stimulation by alternating current with a fixed frequency. In Fig. 2, the fragments of amplified information signal obtained from the sound sensor and fragments of their spectral characteristics are shown.

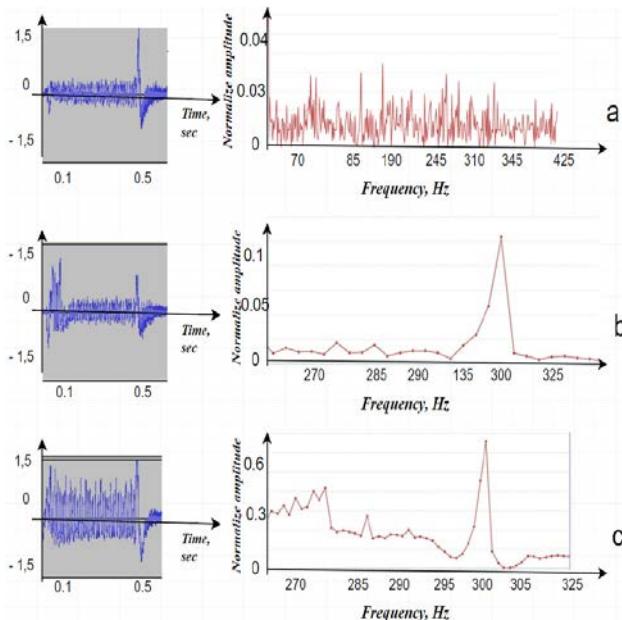


Fig. 2. Result of stimulation of RLN by alternating current with frequency of 300 Hz.

We can see the result of stimulation of the muscle tissue at a distance of more than 1 cm from RLN, with a specific blurred spectrum, without a clearly distinguished

main spectral component in Fig. 2 a). Fig. 2 b) reflects the result of stimulation of the muscle tissue at a distance of not more than 3 mm, with a specific distinguished main spectral component with a small amplitude value. Finally, the result of RLN stimulation with a specific main spectral component with a sufficiently high normalized amplitude (6 times higher than in the previous case) is illustrated in Fig. 2 c).

The described properties give a possibility to develop the method, mathematical model and to improve the RLN identification technology in general.

### III. MATHEMATICAL MODEL FOR RLN IDENTIFICATION

Let's represent obtained set of stimulated points in such form:

$$[z_{i,j}] = [z_{i,j}^-; z_{i,j}^+], i = 1, \dots, I, j = 1, \dots, J, \quad (1)$$

where  $[z_{i,j}]$  is an interval estimation of the normalized amplitude of main spectral component;  $i, j$  are indices of discrete increments of coordinate values on  $X$  and  $Y$  axes relatively to some initially given point. Interval estimation of the amplitude  $[z_{i,j}]$  is caused by the fact that different values of main spectral component amplitude  $z_{i,j}$  may be obtained for equal values of  $i$  and  $j$ . In addition, there is some error of detecting the point with coordinates  $i, j$ .

A mathematical model for RLN identification is considered as a discrete difference model (DDM), that is, the difference scheme in such form [1, 7]:

$$[\hat{v}_{i+1,j+1}] = [\hat{v}_{i+1,j+1}^-; \hat{v}_{i+1,j+1}^+] = \vec{f}^T([\hat{v}_{0,0}], \dots, [\hat{v}_{0,j}], \dots, [\hat{v}_{i,0}], \dots, [\hat{v}_{i,j}] * \vec{g}, \quad i = d + 1, \dots, I, j = d + 1, \dots, J, \quad (2)$$

where  $\vec{f}^T(\bullet)$  is a vector of unknown basis functions that define the structure of DDM;  $\hat{v}_{i,j}$  is a predicted value of main spectral component amplitude in the point with discrete specified spatial coordinates  $i, j$ ;  $\vec{g}$  is a vector of unknown parameters of DDM;  $d$  is order of DDM. Further, the model (2) will be called an interval discrete difference model (IDDM).

To identify this model, the results of measurement of the main spectral component amplitude in the points of stimulation are used.

Based on the requirements of ensuring the accuracy of the model within the accuracy of the experiment, the setting of IDDM (2) will be realized using such criterion [2]:

$$[\hat{v}_{i,j}^-; \hat{v}_{i,j}^+] \subset [\hat{z}_{i,j}^-; \hat{z}_{i,j}^+], \forall i = 1, \dots, I, \forall j = 1, \dots, J. \quad (3)$$

By substituting in the expression (3), the recurrent expression (2) instead of the interval estimations  $[\hat{v}_{i,j}^-; \hat{v}_{i,j}^+]$ , together with the defined initial interval values, we obtain the interval system of non-linear algebraic equations (ISNAE) [12]:

$$\left\{ \begin{array}{l} [\hat{v}_{0,0}^-; \hat{v}_{0,0}^+] \subseteq [\hat{z}_{0,0}^-; \hat{z}_{0,0}^+]; \\ \vdots \\ [\hat{v}_{i-2,j-2}^-; \hat{v}_{i-2,j-2}^+] \subseteq [\hat{z}_{i-2,j-2}^-; \hat{z}_{i-2,j-2}^+]; \\ [\hat{v}_{i-1,j-1}] = \vec{f}^T([\hat{v}_{0,0}], \dots, [\hat{v}_{i-2,j-2}], \vec{u}_o) * \vec{g}; \\ z_{i,j}^- \leq \vec{f}^T([\hat{v}_{0,0}], \dots, [\hat{v}_{i-1,j-1}], \vec{u}_{i,j}, \dots, \vec{u}_{i,j}) * \vec{g} \leq z_{i,j}^+ \\ z_{i+1,j}^- \leq \vec{f}^T([\hat{v}_{0,0}], \dots, [\hat{v}_{i-1,j-1}], \vec{u}_{i,j}, \dots, \vec{u}_{i,j}) * \vec{g} \leq z_{i+1,j}^+ \\ i = 2, \dots, I, j = 2 \end{array} \right.$$

The solution of the obtained system is a vector of unknown parameters of IDDM  $\vec{g}$ . Methods of solving of this system are described in the papers [11-12].

The obtained model in the form of IDDM (2) is graphically represented in Fig. 3.

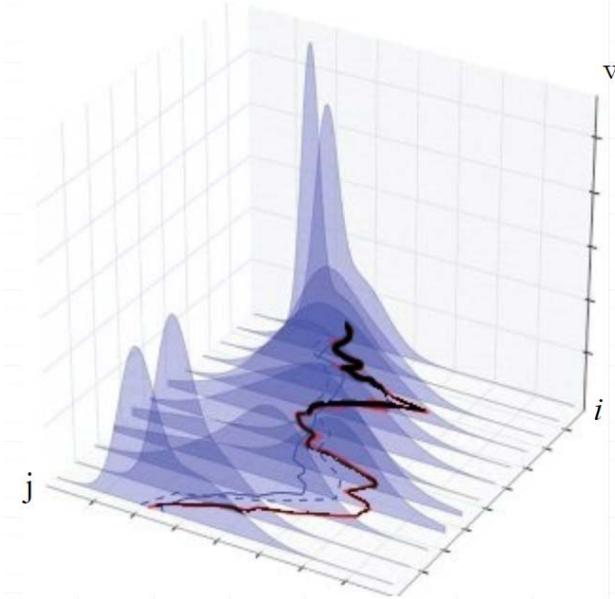


Fig. 3. Visualization of RLN location using IDDM (2).

As we see in Fig. 3, the distribution of values of amplitude of the main spectral component of signals as the reaction on RLN stimulation and surgical wound tissues stimulation is marked with grey color. The RLN location is marked by black colored line on the plane.

#### IV. METHODS AND TOOLS OF ELECTROPHYSIOLOGICAL MONITORING RLN

For RLN location identification in the surgery area during surgery on neck organs, we improved the information technology. This technology consists of 4 main steps, a detailed description of which is given below. The scheme of the information technology is shown in Fig. 4.

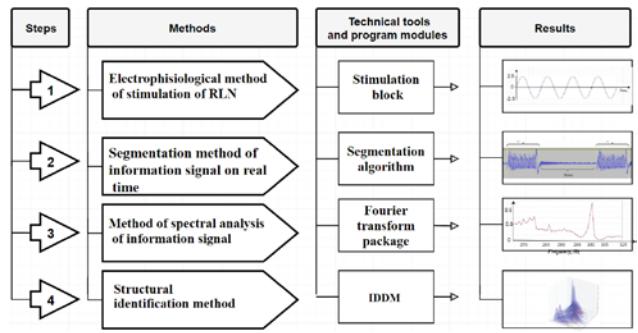


Fig.4. Scheme of information technology for information signal (as a result of stimulation of surgical wound of tissues) processing.

*Step 1. Obtaining of information signal (as result of stimulation of surgery wound tissues).*

At this step, the multifunctional block for surgical wound tissues stimulation is used. The result of stimulation is recorded by the sound sensor. The obtained information signal is digitized and processed by the single-board computer Raspberry Pi 3. The stimulation results are processed in a real-time mode. Also, they may be stored on the external data storages.

*Step 2. Segmentation of information signal.*

This step is needed to highlight the patient reaction on tissues stimulation from sound signal taken during inhale or exhale.

In Fig. 4, the information signal is displayed and the segmentation principle is visually represented. As we can see from the Fig. 2, there are two segments. Each of them is result of the stimulation of RLN. The intervals between the segments represent delay of a patient's breath. These intervals are not informative for tissue classification method and named as the noise.

Unlike existing method of segmentation "synchronization with the appearance of the current of stimulation", main approach of automatic segmentation is based on the principle of threshold choice of an informative segment.

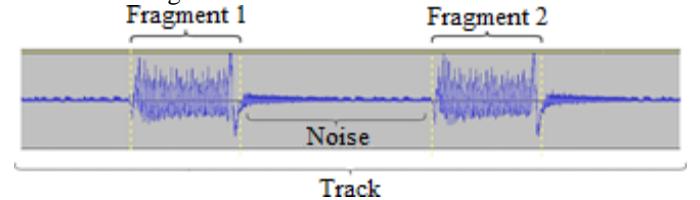


Fig. 5 Illustration of the information signal segmentation.

Because information signal is represented in digital form, for determining of segment beginning to estimate the energy threshold of current,  $n$  countdowns are proposed:

$$E = \sum_{i=1}^n u_i^2, \quad (4)$$

where  $u_i$  is  $i$ -th countdown of information signal.

If this energy exceeds the threshold, then, this is the beginning of the segment:

$$E \geq E_{tr}, \text{ then, } u_{start} = u_n. \quad (5)$$

If the energy of  $n$  counts is less than the threshold, then, this is the end of the segment:

$$E \leq E_{tr}, \text{ then, } u_{stop} = u_n. \quad (6)$$

So, the resulting segment consists of a set of countdowns:

$$U = \{u_i \in [u_{start}; u_{stop}]\}, \quad (7)$$

where  $[u_{start}; u_{stop}]$  is interval of countdowns of determined signal.

#### *Step 3. Selecting of the main spectral component.*

After segmentation of information signal recorded by sound sensor, it is necessary to select the main spectral component of the information signal. At first, the Fourier transform is used for this.

As it is shown in Fig. 2, the amplitude of the main spectral component depends on the distance from the stimulation point to the RLN. Therefore, the amplitude of the selected spectral component is in inverse proportion to this distance.

The obtained at this stage data array is represented in the form (1).

#### *Step 4. Modeling of distribution of the main spectral component amplitudes.*

A mathematical model for RLN identification is considered as an interval discrete dynamic model that is, the difference scheme in form (2). For its identification, the method of structural and parametric identification based on the behavioral model of artificial bee colony [13-14] is used. Behavioral model of artificial bee colony imitates the foraging behavior of the honeybee colony [13].

The application of this IDDM structural identification method involves the implementation of activity phases of all types of bees in the colony: onlooker bees, employed bees and scout bees. Let's consider all stages of implementation of IDDM structural identification method in more details. The result, at this stage, is the model that graphically represented in Fig. 3. Probable RLN location is illustrated by black line.

The test examples with sample of over 1500 stimulation points for different patients showed that using presented technology allow to decrease the RLN damage risk from 20% to 14%. So, the proposed methods and tools for electrophysiological RLN monitoring have a good perspective of development and application.

## VI. CONCLUSION

The proposed methods and tools for electrophysiological RLN monitoring and identification are realized using single-board computer Raspberry Pi 3. Applying of this methods for previously obtained sample of signals (reaction on stimulation of surgical wound tissue) showed good results. In particular, it was shown that for sample of over 1500 stimulation points for different patients using this technology, the risk of RLN damage is decreased from 20% to 14%.

## ACKNOWLEDGMENT

This research was supported by National Grant of Ministry of Education and Science of Ukraine "Mathematical tools and software for classification of

tissues in surgical wound during surgery on the neck organs" (0117U000410).

## REFERENCES

- [1] Abstract book of First World Congress of Neural Monitoring in Thyroid and Parathyroid Surgery, Krakow, Poland, 2015, 161 pp.
- [2] M.C.D. Poveda, G. Dionigi, A. Sitges-Serra, M. Barczynski, P. Angelos, H. Dralle, E. Phelan and G. Randolph. "Intraoperative Monitoring of the Recurrent Laryngeal Nerve during Thyroidectomy: A Standardized Approach (Part 2)". *World Journal of Endocrine Surgery*, vol. 4, no. 1, pp. 33-40, 2012.
- [3] V.K. Dhillon and R.P. Tufano. "The pros and cons to real-time nerve monitoring during recurrent laryngeal nerve dissection: an analysis of the data from a series of thyroidectomy patients". *Gland Surgery*, vol. 6, no. 6, pp. 608-610, 2017.
- [4] H.Y. Kim, X. Liu, C.W. Wu, Y.J. Chai and G. Dionigi. "Future Directions of Neural Monitoring in Thyroid Surgery". *Journal of Endocrine Surgery*, vol. 17, no. 3, pp. 96-103, 2017.
- [5] W.E. Davis, J.L. Rea, J. Templer. "Recurrent laryngeal nerve localization using a microlaryngeal electrode". *Otolaryngology – Head and Neck Surgery*, vol. 87, no. 3, pp. 330-333, 1979.
- [6] M. Dyvak, O. Kozak, A. Pukas. "Interval model for identification of laryngeal nerves". *Przeglad Elektrotechniczny*, vol. 86, no. 1, pp. 139-140, 2010.
- [7] M. Dyvak, N. Kasatkina, A. Pukas, N. Padletska. "Spectral analysis the information signal in the task of identification the recurrent laryngeal nerve in thyroid surgery". *Przeglad Elektrotechniczny*, vol. 89, no. 6, pp. 275- 277, 2013.
- [8] M. Cantelon, M. Harter, T.J. Holowaychuk and N. Rajlich. "Node.js in Action. Shelter Island," NY: Manning Publications, 2013, 416 pp.
- [9] P. Teixeira. "Professional Node.js: Building Javascript Based Scalable Software," Indianapolis, IN: John Wiley & Sons, Inc., 2012, 408 pp.
- [10] Bush, Steve (25 May 2011). "Dongle computer lets kids discover programming on a TV.". *Electronics Weekly*. Retrieved 11 July 2011..
- [11] I. Voytyuk, M. Dyvak, V. Spilchuck, "Research of quality characteristics of models structure in kind of interval difference operator," *11th International Conference The Experience of Designing and Application of CAD Systems in Microelectronics (CADSM)*, pp. 87-87, 2011.
- [12] M. Dyvak, N. Porplytsya, Y. Maslyak, M. Shynkaryk "Method of Parametric Identification for Interval Discrete Dynamic Models and the Computational Scheme of Its Implementation," *Advances in Intelligent Systems and Computing II: Selected Papers from the International Conference on Computer Science and Information Technologies, CSIT 2017*, pp.101- 112, 2018.
- [13] D. Karaboga, B. Gorkemli, C. Ozturk, N. Karaboga, "A comprehensive survey: artificial bee colony (ABC) algorithm and applications," *Artificial Intelligence Review*, vol. 42, no.1, pp. 21-57, 2014.
- [14] D. Karaboga, "An idea based on honey bee swarm for numerical optimization," *Technical Report TR06*, Computer Engineering Department, Engineering Faculty, Erciyes University, Turkey, 2005.

# Modified Method of Subtractive Clustering for Modeling of Distribution of Harmful Vehicles Emission Concentrations

Mykola Dvavak<sup>1</sup>, Yurii Maslyiak<sup>2</sup>, Iryna Voytyuk<sup>3</sup>, Bogdan Maslyiak<sup>4</sup>

Faculty of Computer Information Technologies, Ternopil National Economic University, UKRAINE, Ternopil, 8 A. Chekhova str., email:  
[mdy@tneu.edu.ua](mailto:mdy@tneu.edu.ua)<sup>1</sup>, [yuramasua@gmail.com](mailto:yuramasua@gmail.com)<sup>2</sup>, [i.voytyuk@tneu.edu.ua](mailto:i.voytyuk@tneu.edu.ua)<sup>3</sup>, [bm@tneu.edu.ua](mailto:bm@tneu.edu.ua)<sup>4</sup>

**Abstract:** Mathematical modeling of distribution of harmful vehicle emissions concentrations is considered in the paper. The modified subtractive clustering method for modeling is proposed. This method is characterized by its implementation simplicity due to the fact that it does not require a large sample of experimental data and does not require to set a predetermined number of clusters. An example of clustering method application for data preparation for modeling of distribution of harmful vehicle emissions concentrations is given.

**Keywords:** mathematical modeling, clustering analysis, interval analysis, interval difference operator, harmful vehicle emissions, vehicular traffic.

## I. INTRODUCTION

One of the biggest problems of large and medium cities is the pollution of the surface atmospheric layer and soils by harmful emissions from vehicles. Large amount of harmful substances is concentrated in the vehicles exhaust gases. Among them, with high concentrations: Nitrogen oxides, Carbon oxides and Sulfur oxides. Motor transport is a source of atmospheric and soils pollution, which should be considered as distributed one. For reflecting and predicting of concentrations of harmful vehicle emissions, it is expedient to use mathematical models. They can be built based on the results of selective observations of their dynamics with known boundary errors of measurements. This approach was considered in [1].

The process of harmful emissions distribution and its dynamics is considered as a mass transfer process. For its description, the difference operators (schemes) are used. Their identification is carried out using the data of measurements of harmful emissions concentrations with known boundary errors. Such data are called interval data [2,3]. As is known, the methods of difference operators identification based on the interval data analysis require a uniform measurement grid that is impossible for the real city conditions. Mostly, the measurements of harmful emission concentrations is carried out in places with intensive traffic and accumulation of vehicles. This means that measurement grid is not uniform. Thus, for building of mentioned models, it is necessary to solve three tasks related to data preparation: execute cluster analysis for defining of homogeneous vehicular traffic intensity areas; identify the discrete values of the grid step; calculate estimates of harmful vehicle emissions concentrations in the nodes of the grid. The third

task is solved by methods of interpolation [4,5]. The first one and second one are the subjects for research of this work.

## II. STATEMENT OF THE PROBLEM

To solve the environmental monitoring tasks, it is necessary to build models of stationary and non-stationary fields of concentrations of harmful vehicle emissions [6]. The theoretical basis for solving this type of tasks are the mathematical models of objects with distributed parameters in the form of partial differential equations. Concentration of attention on the physical properties of environment requires to significantly complicate the mathematical model. Even despite the fact that, in practice, it is impossible to verify the results of modeling with real data obtained under conditions that meet the conditions of modeling. First of all, this is related to the complexity of the measurement experiment. For example, if a mathematical model in the form of a differential equation accurately enough describes the process of transferring of chemical substances in the atmosphere in case of wind gusts or other turbulent phenomena in the atmosphere, then an integrated value of the chemical substance concentration per volume unit is established in the process of measurement. In addition, the accuracy of such measurements is low, the relative measurement error may reach 50%. Consequently, it is enough to build a mathematical model with an accuracy that is equivalent to the accuracy of the measurement experiment. At this, it is expedient to represent the experimental data in the form of intervals of possible values of the modeled characteristic:

$$[z_{i,j,h,k}^-; z_{i,j,h,k}^+], \quad i=1, \dots, I, j=1, \dots, J, h=1, \dots, H, k=1, \dots, K \quad (1)$$

where  $z_{i,j,h,k}^-$ ,  $z_{i,j,h,k}^+$  are the lower and upper bounds of the interval of possible values of measured concentration of harmful substances in the grid nodes with discretely given spatial coordinates  $i=1, \dots, I$ ,  $j=1, \dots, J$ ,  $h=1, \dots, H$  at the discrete time value  $k=1, \dots, K$ , respectively.

It is worth to note that, in the measurement experiment, we can set the lower and upper bounds based on the relative error of the measuring device:  $z_{i,j,h,k}^- = z_{i,j,h,k} - z_{i,j,h,k} \cdot \varepsilon$  and  $z_{i,j,h,k}^+ = z_{i,j,h,k} + z_{i,j,h,k} \cdot \varepsilon$ , where  $z_{i,j,h,k}$  is the measured value of the harmful substance concentration;  $\varepsilon$  is relative measurement error.

Under these conditions, macromodeling is the only way to reflect the distribution of harmful emissions concentrations. The building of such macromodels is convenient to carry out based on the obtained interval data in the form (1).

In the papers of O.G. Ivakhnenko [7], the inductive approach is described for choosing of acceptable way of mathematical description of these processes. Its essence consists in defining of some difference scheme in the way of its adjustment in accordance with the experimental data. The difference scheme itself, which actually converts the values of input variables into output values, is called a difference operator. The process of adjustment of this scheme is called structural identification [8,12].

In general case, the expression of linear in parameters difference operator (DO) has the following form [2]:

$$\begin{aligned} v_{i,j,h,k} = \vec{f}^T(v_{0,0,0,0}, \dots, v_{0,0,h-1,0}, v_{i-1,0,0,0}, \dots, v_{0,j-1,0,0}, \dots, \\ v_{i-1,j-1,h-1,k-1}, \vec{u}_{i,j,h,0}, \dots, \vec{u}_{i,j,h,k}) \cdot \vec{g}, \quad (2) \end{aligned}$$

$$i=d, \dots, I, j=d, \dots, J, h=d, \dots, H, k=d, \dots, K$$

where  $\vec{f}^T(\bullet)$  is the vector of basis functions (nonlinear, in general case) by which, the transformation of the modeled characteristic values, as well as the input variables in the spatial grid nodes for the certain discrete moments of time is carried out;  $v_{i,j,h,k}$  modeled concentration of harmful emissions in grid nodes with discretely-given spatial coordinates  $i=d, \dots, I$ ,  $j=d, \dots, J$ ,  $h=d, \dots, H$  at the moments of time  $k=d, \dots, K$ ;  $\vec{u}_{i,j,h,0}, \dots, \vec{u}_{i,j,h,k}$  are the vectors of input variables (controls);  $d$  is the DO order;  $\vec{g}$  is the vector of unknown parameters of DO.

As a result of executing of structural identification procedure, we establish the difference computational scheme, in particular: the basis functions vector  $\vec{f}^T(\bullet)$ ; sets and dimensionality of input variables (controls) vectors  $\vec{u}_{i,j,h,0}, \dots, \vec{u}_{i,j,h,k}$ ; order of the difference scheme  $d$ , which, as is known, is equivalent to the order of the differential equation (the analogue of the difference scheme). To realize the difference scheme, it is also necessary to set the initial conditions, that is, the value of each discrete element from the set  $v_{0,0,0,0}, \dots, v_{0,0,h-1,0}, v_{i-1,0,0,0}, \dots, v_{0,j-1,0,0}, \dots, v_{i-1,j-1,h-1,k-1}, \vec{u}_{i,j,h,0}, \dots, \vec{u}_{i,j,h,k}$  (as a rule, the initial ones) and to establish the values of the parameters vector  $\vec{g}$  components. If the structure of DO is known then, it remains the actual task of adjusting the parameters of DO (2) in such a way as to ensure the maximal consistency between the modeled characteristic and experimentally obtained values of this characteristic. Such a task is called the parametric identification task [9,13].

Based on the requirements of ensuring the mathematical model accuracy within the bounds of the measuring experiment accuracy, the conditions of consistency between experimental data, represented in the interval form (1), and data obtained based on the mathematical model in the form of DO (2), can be formulated in such form:

$$v_{i,j,h,k} \subset [z_{i,j,h,k}^-; z_{i,j,h,k}^+], \quad (3)$$

$$\forall i=d, \dots, I, \forall j=1, \dots, J, \forall h=d, \dots, H, \forall k=d, \dots, K$$

Based on the results of conducted analysis, we can state that for ensuring of conditions of the given accuracy (3) of the macromodel in the form of linear DO (2) during solving the task of its parametric identification, the application of interval data analysis methods [9] is substantiated.

Let's assume that the vector of parameters estimates  $\hat{\vec{g}}$  in the DO (2) is obtained based on the interval data analysis. Substituting the vector of parameters estimates  $\hat{\vec{g}}$  of DO instead of their true values  $\vec{g}$  in expression (2) together with given initial interval values of each element in the set  $[\hat{v}_{0,0,0,0}], \dots, [\hat{v}_{0,0,h-1,0}], [\hat{v}_{i-1,0,0,0}], \dots, [\hat{v}_{0,j-1,0,0}], \dots, [\hat{v}_{i-1,j-1,h-1,k-1}]$  and given vectors of the input variables  $\vec{u}_{i,j,h,0}, \dots, \vec{u}_{i,j,h,k}$ , we obtain an interval estimate of the harmful substance concentration  $[\hat{v}_{i,j,h,k}]$  in the nodes with discretely given spatial coordinates  $i=1, \dots, I$ ,  $j=1, \dots, J$ ,  $h=1, \dots, H$  at discrete moments of time  $k=1, \dots, K$ :

$$\begin{aligned} [\hat{v}_{i,j,h,k}] = [\hat{v}_{i,j,h,k}^-; \hat{v}_{i,j,h,k}^+] = \vec{f}^T([\hat{v}_{0,0,0,0}], \dots, [\hat{v}_{0,0,h-1,0}], \\ [\hat{v}_{i-1,0,0,0}], \dots, [\hat{v}_{0,j-1,0,0}], \dots, [\hat{v}_{i-1,j-1,h-1,k-1}], \\ \vec{u}_{i,j,h,0}, \dots, \vec{u}_{i,j,h,k}) \cdot \hat{\vec{g}}, \quad (4) \end{aligned}$$

$$i=1, \dots, I, j=1, \dots, J, h=1, \dots, H, k=1, \dots, K$$

Thus, the mathematical model of stationary and non-stationary fields of harmful emissions concentrations for the task of environmental control will be described by a DO in general form (4). Taking into account that all calculations in equation (4) are carried out using interval arithmetic rules [2], the difference operator (4) is called an interval difference operator (IDO).

The conditions of consistency of experimental data, represented in interval form (1), with the data obtained based on macromodel in the form of IDO (4) are formulated as follows:

$$[\hat{v}_{i,j,h,k}^-; \hat{v}_{i,j,h,k}^+] \subset [z_{i,j,h,k}^-; z_{i,j,h,k}^+], \quad (5)$$

$$\forall i=1, \dots, I, \forall j=1, \dots, J, \forall h=1, \dots, H, \forall k=1, \dots, K$$

Let's substitute in expressions (5), instead of interval estimates of the harmful substance concentrations  $[\hat{v}_{i,j,h,k}^-; \hat{v}_{i,j,h,k}^+]$ , its interval values calculated using IDO (4), together with taking into account the given initial interval values of each element from the set

$$[\hat{v}_{0,0,0,0}] \subseteq [z_{0,0,0,0}], \dots, [\hat{v}_{0,0,h-1,0}] \subseteq [z_{0,0,h-1,0}],$$

$$[\hat{v}_{i-1,0,0,0}] \subseteq [z_{i-1,0,0,0}], \dots, [\hat{v}_{0,j-1,0,0}] \subseteq [z_{0,j-1,0,0}], \dots, (6)$$

$$[\hat{v}_{i-1,j-1,h-1,k-1}] \subseteq [z_{i-1,j-1,h-1,k-1}]$$

and given vectors of input variables  $\vec{u}_{i,j,h,0}, \dots, \vec{u}_{i,j,h,k}$ . We obtain such interval system of non-linear algebraic equations (ISNAE) [3]:

$$\left\{ \begin{array}{l} [\hat{v}_{0,0,0,0}^-, \hat{v}_{0,0,0,0}^+] \subseteq [z_{0,0,0,0}^-, z_{0,0,0,0}^+], \dots, \\ [\hat{v}_{i-d,j-d,h-d,k-d}^-, \hat{v}_{i-d,j-d,h-d,k-d}^+] \subseteq [z_{i-d,j-d,h-d,k-d}^-, z_{i-d,j-d,h-d,k-d}^+]; \\ [\hat{v}_{i-1,j-1,h-1,k-1}] = \vec{f}^T([\hat{v}_{0,0,0,0}], \dots, [\hat{v}_{0,0,h-1,0}], [\hat{v}_{i-1,0,0,0}], \dots, \\ [\hat{v}_{0,j-1,0,0}], \dots, [\hat{v}_{i-d,j-d,h-d,k-d}], \vec{u}_0, \dots, \vec{u}_{k-1}) \cdot \hat{\vec{g}}; \\ z_{i,j,h,k}^- \leq \vec{f}^T([\hat{v}_{0,0,0,0}], \dots, [\hat{v}_{0,0,h-1,0}], [\hat{v}_{i-1,0,0,0}], \dots, [\hat{v}_{0,j-1,0,0}], \dots, \\ [\hat{v}_{i-d,j-d,h-d,k-d}], \vec{u}_0, \dots, \vec{u}_k) \cdot \hat{\vec{g}} \leq z_{i,j,h,k}^+; \\ i = d, \dots, I, j = 2, \dots, J, h = d, \dots, H, k = d, \dots, K. \end{array} \right. \quad (7)$$

So, the ISNAE (7) is obtained by substituting the interval estimates of the output characteristic (given in the form of initial conditions and predicted using expression (4) in the remaining nodes of the grid) into conditions (5). Therefore, the task of parametric identification of IDO (4) under conditions (5) is the task of solving ISNAE in the form (7). Methods for estimation of solutions of the obtained ISNAE are described in [10].

The analysis of the proposed scheme for building of mathematical model of harmful vehicle emissions distribution showed that before its implementing, it is necessary to obtain a uniform grid of measured concentrations (3) in its nodes and vectors of influences on them  $\vec{u}_{i,j,h,0}, \dots, \vec{u}_{i,j,h,k}$ . The main among them, is the vehicular traffic intensity. This task is solved using modified method of subtractive clustering of data [11] on the traffic intensity.

### III. MODIFIED SUBTRACTIVE CLUSTERING METHOD

As the basis for method of clustering of vehicular traffic distribution, it is expedient to use the “mountain” clustering method with subtractive algorithm of its implementation. This method does not require a large sample of experimental data and does not require to set a predetermined number of clusters that significantly reduces the time for its implementation. It is also worth to note that the number of clusters based on this method is regulated by the only parameter which is the cluster radius [11].

According to the clustering method, in the beginning, we form the potential cluster centers from the rows of data matrix for the clustering of input variables and calculate the potentials of identified cluster centers using the expression:

$$P_h(c_h) = \sum_{k=1}^K \exp(-\alpha \cdot \|\vec{c}_h - \vec{x}_k\|), \quad (8)$$

where  $\vec{c}_h = (c_{1h}, c_{2h}, \dots, c_{Kh})$  is a potential center of  $h$ -th cluster;  $\alpha$  is a positive constant;  $\|\vec{c}_h - \vec{x}_k\|$  is a distance between potential center of  $h$ -th cluster  $\vec{c}_h$  and input experimental data  $\vec{x}_k$ ,  $k=1, \dots, K$ ,  $h=1, \dots, H$ ;  $H$  is a number of possible clusters.

In our case, if the only property of a cluster which is the number of vehicles  $u_{xi,yj,k}$  in the point with coordinates  $x_i, y_j$  at a discrete moment of time  $k$  is taken into account, the expression for estimation of potentials of given cluster centers, has such form:

$$P_h(x_h, y_h, k) = \sum_{i=1}^I \sum_{j=1}^J \exp\left(-\alpha \cdot \|\bar{u}_{x_h, y_h, k} - u_{x_i, y_j, k}\|\right), \quad (9)$$

where  $P_h(x_h, y_h, k)$  is potential of a point (center of cluster with coordinates  $x_h, y_h$  at moment of time  $k$ );  $\bar{u}_{x_h, y_h, k}$ ,  $u_{x_i, y_j, k}$  are the numbers of motor vehicles in a point of potential cluster center  $x_h, y_h, k$  and in points  $x_i, y_j, k$  with defined traffic intensity and measured concentrations, respectively.

The illustration of the potentials distribution is represented as a surface in the form of a mountainous relief (Fig. 1), whose peaks have the highest potential values and are pretenders to be the centers of the formed clusters.

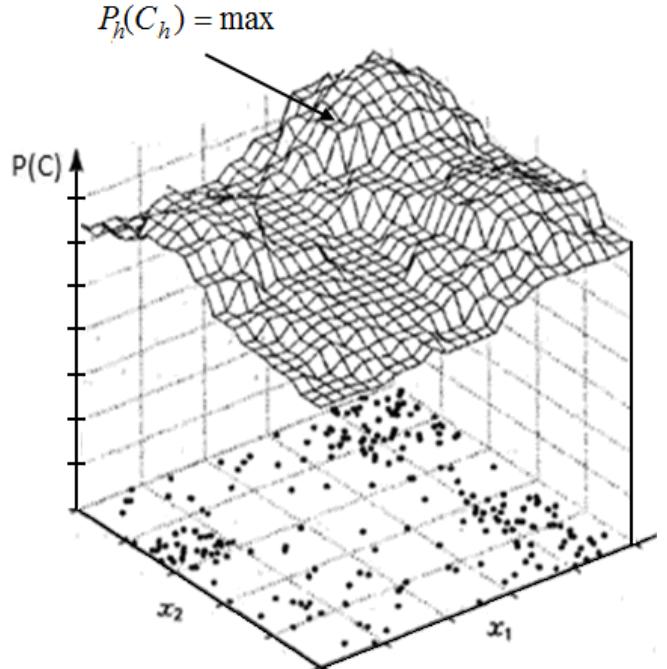


Fig. 1. The illustration of potentials distribution based on the mountain clustering method.

As we can see in the Fig. 1, one “mountain peak” is surrounded by other peaks that causes the problem of building of very similar data clusters with the corresponding centers. This does not provide the high quality clustering results.

As centers of the clusters, we choose the coordinates of “mountain peaks”, that is, the center of the cluster is the point on the city map with the highest value of potential:

$$(x_h, y_h, k) = \arg \max_{h=1, \dots, H} P_h(x_h, y_h, k). \quad (10)$$

In order to avoid the building of similar clusters, we must recalculate the potential values for the remaining possible cluster centers:

$$\begin{aligned} P_{h+1}(x_{h+1}, y_{h+1}, k) &= P_{h+1}(x_h, y_h, k) - P_h(x_h, y_h, k) \cdot \\ &\exp(-\beta \cdot \|\bar{u}_{x_{h+1}, y_{h+1}, k} - \bar{u}_{x_h, y_h, k}\|), h = 1, \dots, H, \end{aligned} \quad (11)$$

where  $P_h(x_h, y_h, k)$  is a potential center of  $h$ -th cluster on  $h$ -

th iteration;  $P_{h+1}(x_h, y_h, k)$  is a potential of center of  $h$ -th cluster on  $h+1$  iteration;  $\beta$  is a positive constant,  $\|\bar{u}_{x_{h+1}, y_{h+1}, k} - \bar{u}_{x_h, y_h, k}\|$  is a distance between potential center of  $h+1$  cluster and center of found  $h$ -th cluster.

The procedure of cluster centers calculation is carried out until all the rows of the input variable matrix  $X$ , which is represented by the set (3), are excluded.

The above procedure is based on the subtractive clustering algorithm, which is based on the following steps.

*Step 1.* Forming of potential cluster centers. They are all points of measured harmful emission concentrations and corresponding intensities of vehicular traffic.

*Step 2.* Calculation the potential of possible cluster centers based on (9).

*Step 3.* Selecting the data point with the maximal potential for representation of the cluster center based on (10).

*Step 4.* Excluding the influence of the found cluster center in the way of recalculating the potentials for other possible cluster centers by (11).

*Step 5.* Identifying the next cluster and the coordinates of its center. If the maximal value of the cluster center potential exceeds some predetermined threshold which is the cluster radius, that is  $P_h(x_h, y_h, k)$ , then proceed to *step 4*, otherwise, the algorithm is completed.

The iterative procedure for identification of cluster centers and the recalculation of potentials is repeated until all points in the space of input experimental data are located within the neighborhoods of the radius of sought cluster centers.

As a result of the clustering algorithm implementation, we obtain  $h$  clusters,  $h = 1, \dots, H$ , with the corresponding centers. The next step is the identification of a uniform grid nodes for homogeneous parts of a cluster. The discrete values of the grid nodes coordinates are equal to the cluster diameter, and the value  $\bar{u}_{x_h, y_h, k}$  is the number of vehicles in the point of a cluster center  $x_h, y_h, k$ . To assign the number of vehicles at  $k$ -th moment to the grid nodes, it is enough to analyze, what cluster the node is in. If the node belongs to the  $h$ -th cluster, then, the number of vehicles in the node is  $\bar{u}_{x_h, y_h, k}$ .

#### IV. EXAMPLE OF CLUSTERING METHOD APPLICATION

Let's consider the application of the developed clustering method for obtaining of uniform grid of nodes on an example of Ternopil city.

The fragment of map of central part of Ternopil city with the marked points of the measured vehicular traffic intensity for one discrete time (one hour) is shown in the Fig. 2. As we can see, the traffic is distributed not uniformly over the territory. Therefore, it is advisable to measure its intensity at some selected points, where this intensity is the highest, for example, as it is shown on the map of Ternopil city.

The points of measurement of traffic intensity are colored red on the map.

The application of cluster analysis for determination of areas with specific vehicular traffic intensities under condition of identification of cluster centers that are located on a certain uniform grid gives a possibility to define the

discretization step for building a DO. In our case, the cluster is the set of points of a certain area of the city with similar values of current vehicles number.

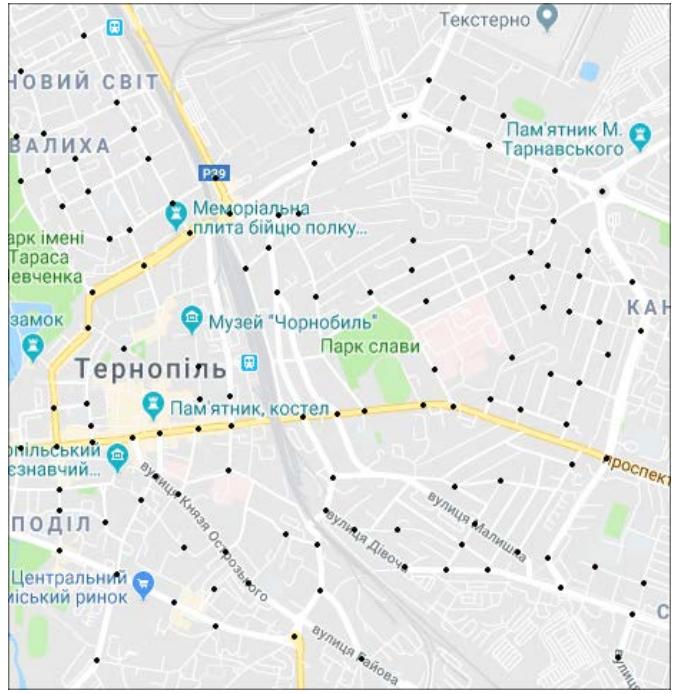


Fig. 2. Points of measurements of vehicular traffic intensity on the example of Ternopil city.

The result of the proposed method of cluster analysis application is schematically shown in Fig. 3. As we can see, during the clustering process,  $H$  clusters with different vehicular traffic intensity and radius  $r$  were defined. So, the discrete values of grid nodes coordinates are equal to the cluster diameter.

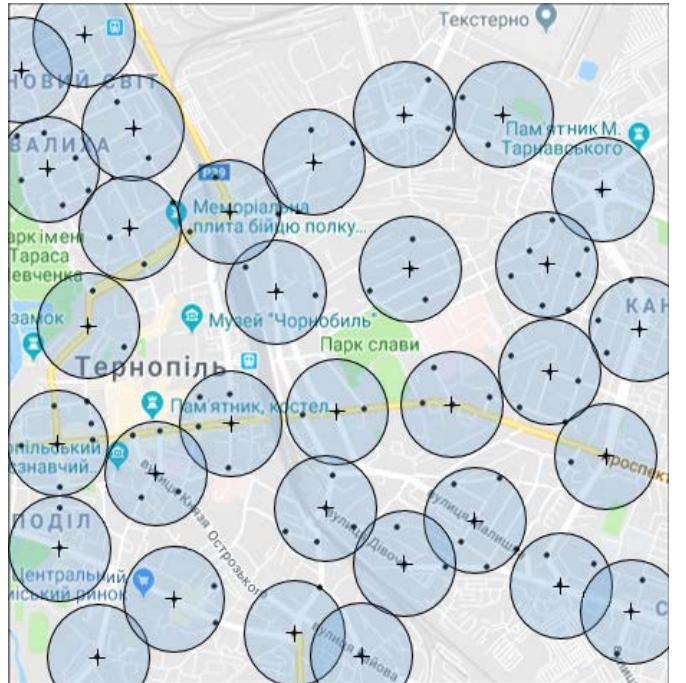


Fig. 3. The result of clustering of vehicles quantity distribution on the Ternopil city map.

Obtained grid for building of distribution model of harmful vehicle emission concentrations in the form of IDO (4) is schematically shown in Fig. 4.

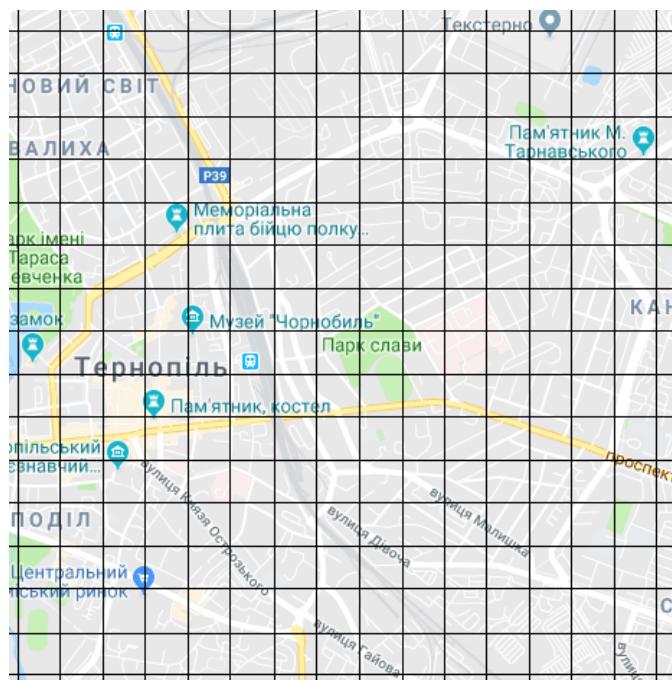


Fig. 4. The uniform grid with known spatially distributed vehicular traffic intensities.

For building of mentioned model in the form of IDO (4), it is enough to execute interpolation and identify the pollution concentrations in the grid nodes.

## V. CONCLUSIONS

The modified method of subtractive clustering and interval analysis for modeling of distribution of harmful vehicle emissions concentrations and vehicular traffic intensity under conditions of non-uniform measurement grid were proposed and substantiated.

## ACKNOWLEDGMENT

This research has been supported by National Grants of Ministry of Education and Science of Ukraine "Mathematical tools and software for control the air pollution from vehicles" (0116U005507) and "Mathematical tools and software for classification of tissues in surgical wound during surgery on the neck organs" (0117U000410).

## REFERENCES

- [1] A. Veremchuk, A. Pukas, I. Voytyuk and I. Spivak, "Mathematical and software tools for modeling objects with distributed parameters," *2016 13th International Conference on Modern Problems of Radio Engineering, Telecommunications and Computer Science (TCSET)*, Lviv, 2016, pp. 149-152.
- [2] I. Voytyuk, M. Dyvak, V. Nemish "Method and genetic algorithm for structure identification of interval differentice operators in the tasks of environmental monitoring", *Collected Works of Donetsk National Technical University Series "Information, cybernetics and computer science."*, 2011, Vol. 14 (188), pp. 8-17.
- [3] M. Dyvak "Mathematical Modeling Tasks of Static Systems with Interval Data", Ternopil: TNEU Publishing House "Economic Thought", 2011, 216 p.
- [4] M. Dyvak, I. Voytyuk, T. Dyvak, A. Pukas "Application of the interval difference operator for approximation of fields of harmful emissions concentration from vehicles", *Measuring and Computing Devices in Technological Processes*, 2011, No. 1 (37), pp. 44-52.
- [5] Kvyetnyy R. N., Dementiev V. Yu., Mashnitsky M. O., Judin O. O. "Difference methods and splines in multidimensional interpolation problems", Vinnitsa: UNIVERSUM, 2009, 87 p.
- [6] N. Ocheretnyuk, I. Voytyuk, M. Dyvak and Y. Martsenyuk, "Features of structure identification the macromodels for nonstationary fields of air pollutions from vehicles," *Proceedings of International Conference on Modern Problem of Radio Engineering, Telecommunications and Computer Science*, Lviv-Slavsk, 2012, pp. 444-444.
- [7] A.G. Ivakhnenko "Inductive method of self-organizing of models of complex systems", Kyiv: Scientific thought, 1981, 296 p.
- [8] M. Dyvak, I. Voytyuk, T. Dyvak, A. Pukas "Application of the interval difference operator for approximation of fields of harmful emissions concentration from vehicles", *Measuring and Computing Devices in Technological Processes*, 2011, No. 34 (110), pp. 86-94.
- [9] T. Dyvak "Parametric identification of interval difference operator on the example of macromodel for distribution of humidity in the drywall sheets in the process of drying", *Information Technologies and Computer Engineering: international Scientific Journal*, 2012, Vol. 3, pp. 79-85.
- [10] M. Dyvak, N. Porplytsya, Y. Maslyak, M. Shynkaryk "Method of Parametric Identification for Interval Discrete Dynamic Models and the Computational Scheme of Its Implementation," *Advances in Intelligent Systems and Computing II: Selected Papers from the International Conference on Computer Science and Information Technologies, CSIT 2017*, pp.101- 112, 2018.
- [11] Shtovba S. Introduction to the theory of fuzzy sets and fuzzy logic. Access mode: <http://matlab.exponenta.ru/fuzzylogic/book1/index.php>
- [12] Porplytsya, N., Dyvak, M., Dyvak, T., Voytyuk, I. "Structure identification of interval difference operator for control the production process of drywall." *Proceedings of 12th International Conference on the Experience of Designing and Application of CAD Systems in Microelectronics, CADSM'2013*, pp. 262-264 (2013).
- [13] Fliess, M., Sira-Ramirez, H. "Closed-loop parametric identification for continuous-time linear systems via new algebraic techniques." *H. Garnier & L. Wang. Identification of Continuous-time Models from sampled Data*, Springer, pp. 362-391, 2008.

# Two-Dimensional Spectral Detector for Baggage Inspection X-Ray System

Mykola Shutko<sup>1</sup>, Volodymyr Shutko<sup>2</sup>, Lidiya Tereshchenko<sup>3</sup>, Maksym Zaliskyi<sup>4</sup>, Iuliia Silantieva<sup>5</sup>

1. Information Security Devices Department, National Aviation University, UKRAINE, Kyiv, 1 Komarova av., email:

[Lenochka597@ukr.net](mailto:Lenochka597@ukr.net)

2. Electronics Department, National Aviation University, UKRAINE, Kyiv, 1 Komarova av., email: [vnshutko@ukr.net](mailto:vnshutko@ukr.net)

3. Aviation Radioelectronic Complexes Department, National Aviation University, UKRAINE, Kyiv, 1 Komarova av., email: [10118@ukr.net](mailto:10118@ukr.net)

4. Aviation Radioelectronic Complexes Department, National Aviation University, UKRAINE, Kyiv, 1 Komarova av., email: [maximus2812@ukr.net](mailto:maximus2812@ukr.net)

5. International Transportation and Customs Control Department, National Transport University, UKRAINE, Kyiv, 1, M. Omelianovycha Pavlenka Str., email: [gmelanine@gmail.com](mailto:gmelanine@gmail.com)

**Abstract:** The paper deals with two-dimensional spectral detector for baggage inspection X-ray devices. This detector is based on construction of analytical models for internal structure of object under control and their spectrum calculation. The methods of projective geometry and Bouguer-Lambert law are applied to obtain the analytical models for shadows of the three-dimensional objects. Spectral detector are designed according to Neyman-Pearson criterion. Analysis shows that proposed spectral detector has good operating characteristics even at low signal-to-noise ratios.

**Keywords:** aviation security service, X-ray, optical imaging, shadow of the three-dimensional objects, spectral detector.

## I. INTRODUCTION

Ensuring effective protection against terrorism is the most difficult issue, especially for countries with a developed air transport network, a large number of airlines and airports. The problem is complicated by unpredictability of terrorists' actions. In addition, vulnerabilities in aviation security systems (such as procedures for screening airline passengers and their baggage, freight shipments, mail, etc.) that can be exploited by law violators should be taken into consideration.

The main way to improve aviation safety is to prevent hazardous objects and substances, explosive devices and weapons on aircraft board. This requires a comprehensive development and introduction of new methods of screening, detection and identification of dangerous objects under control.

Insights of the direct visualization methods indicate that they are inherent in the same type of operations: primary radiation exposure of the objects under control in configuration space (in the case of active method), reradiation reception (scattered or passed through the object), its conversion into an electrical signal, signal processing and electrical-to-optical signal conversion.

## II. LITERATURE REVIEW AND PROBLEM STATEMENT

The paper addresses applied research challenges concerning development and application of a new method of determination (visualization) of the internal structure of the objects under control (OC), that enables dangerous OC to be identified with high probability in real time, increases the speed of dangerous substances identification in luggage, and provides automation of these processes. In addition, automatic generation of images of hazardous OC allows for periodic inspections of aviation security service operators.

Detection systems based on X-ray, computer tomography and spectroscopy of mobile ions have certain shortcomings [1 – 7]. Some of these systems can detect well-hidden explosives, but their implementation requires considerable funds. In addition, they have a high level of false alarms (approximately 0.2 ... 0.4).

Thus, the development of analytical models for the receipt of multidimensional shadows of translucent objects for further processing will allow the classification of OC, which will greatly facilitate the work of operators serving supervision devices in Aviation Security Service (AvSS), reducing the value of false alarms.

Literature analysis showed that modernization of equipment for AvSS is carried out in two directions: in the part of the improvement of hardware and software. In [8] authors proposed new X-ray backscatter technique using an un-collimated powerful (high kW) X-ray beam and an efficient pinhole camera encompassed with a high resolution matrix detector for imaging of an object. Moreover, a high-energy X-ray inspection technique for the reliable inspection of air freight container was presented in [9].

Analysis of various strategies for object detection in X-ray security imagery is given in [10]. Moreover, paper [11] also deals with a technique for the classification of X-ray baggage images using convolutional neural networks. Application of deep convolutional neural network as classification method in medicine X-ray image analysis was considered in [12].

In [13] authors investigated the feasibility of applying straight-line-trajectory-based tomographic imaging configurations to security inspections. The method of automated target recognition with usage of reference database, which contains X-ray images of OC, for cargo scanning systems was proposed in [14]. The papers [15, 16] deal with procedures of handguns, shuriken and razor blades recognition for baggage inspection.

The simulation of the internal structure for OC with simple and complex forms using the point source of irradiation in the center, as well as with the bias relative to the center, is considered in [17]. The method developed for optical imaging of the inner structure of the three-dimensional objects allows obtaining a shadow of these objects, exposed to electromagnetic radiation. It has useful applications in different life spheres, as in medicine, manufacturing industry, in a process of customs supervision of goods and means of transport for commercial use, etc. It allows the AvSS to increase the probability of correct detection of hazardous materials and reduce false alarms of its security system. For medicine the method may help to increase the probability of health hazard anomaly detection.

So aim of this paper is synthesis of two-dimensional spectral detector for baggage inspection X-ray devices.

### III. TWO-DIMENSIONAL SPECTRAL DETECTOR

The construction of an analytical model reduces to the calculation of a projective image of isotropic object in the case of homogeneous irradiation by a point source located on the axis of object symmetry perpendicular to the plane of the image (screen).

To determine a position of the radiation source, the OC and the screen with a point source it is appropriate to use cylindrical coordinate system applied to the Fig. 1. The OC model with complex form is presented in Fig. 2.

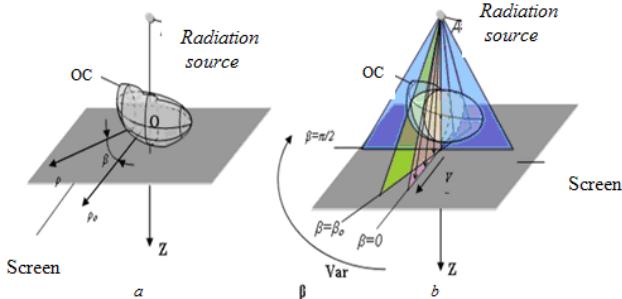


Fig. 1. OC scanning: (a) is the setting a cylindrical coordinate system; (b) is the setting a scanning beam position

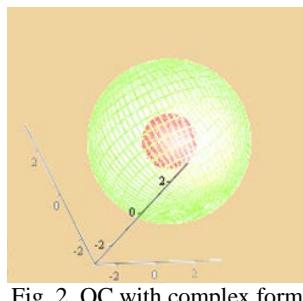


Fig. 2. OC with complex form

Internal visualization of the OC with a complex form, in this case a sphere in the sphere, designed with point source is shown in the Fig. 3.

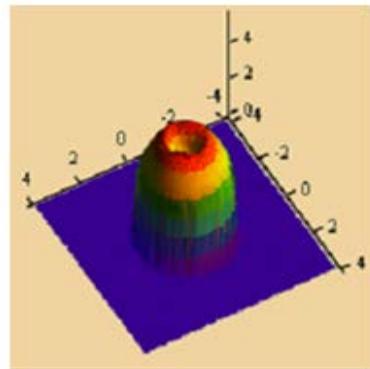


Fig. 3. Inner structure imaging for OC

The simulation shows that the simplest objects have shadows with transient characteristics, half-dooms, distortions of the type of the crater, where there are generally flat irradiating planes. Changing the irradiation angle changes the shadow to unrecognizability. To accurately identify the intended OC, it is necessary to automate the process of recognizing shadows, taking into account possible distances between the source, the OC and the screen-receiver, the irradiation angles, etc.

Methods of analytical modeling of the OC with different shapes, geometrical dimension, foreshortening, substance and appropriate extinction coefficients, used to develop procedures for identifying dangerous objects under security supervision of passengers and baggage, allow to image OC inner structure.

In order to verify the developed models multidimensional spectra of visualization images are obtained.

Procedure for image processing consists of using a shadow of the object of given shape to construct a two-dimensional spectrum and its subsequent use in developing the standard spectral detector proposed in the research. This detector is invariant with respect to the location of the OC in the working area.

The invariance of the calculated spectrum to the location of the OC on the plane of the screen provides the possibility of applying algorithms for the calculation of two-dimensional spatial spectra of the visualization image in relation to the wanted images of some image anomalies in the introscopic imaging systems of the AvSS.

That is, the desired density distribution of the object of control  $\mu(x, y)$  must be matched to fit its two-dimensional spatial spectrum – Fourier-image  $M(k_x, k_y)$ . In the further processing of visualization data, we find solutions in the frequency space  $M(k_x, k_y)$ , and then, through the inverse Fourier transform, the desired distribution is calculated  $\mu^*(x, y)$ . The resulting distribution is selected according to those images, which are in the memory of the supervision system. A decision is made to detect a particular object after matching the resulting image  $\mu^*(x, y)$  and mask  $\mu^*(x, y)$ .

Figures 4, 5 shows the spectra of images of different shades of opaque OC of a simple shape on size a 100x100 screen plane located almost above the center of the screen.

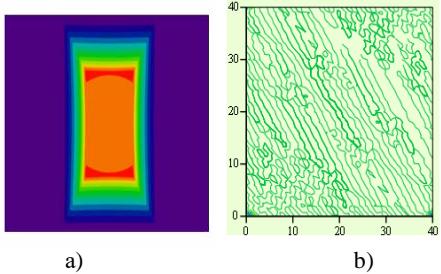


Fig. 4. Shadow (a) of a parallelepiped and its spectrum (b)

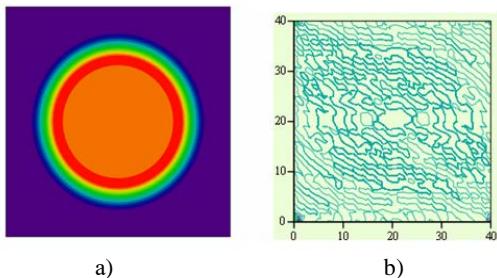


Fig. 5. Shadow (a) of the cylinder and its spectrum (b)

On one plane, the shadows of two parallelepipeds are located, and their spectral images are obtained (Fig. 6).

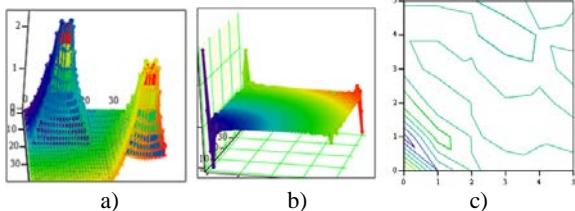


Fig. 6. Shadows of two parallelepipeds and their spectrum:  
a) shadows of two parallelepipeds; b) three-dimensional image of the spectrum those shadows; c) a two-dimensional projection of the spectrum of shadows of parallelepipeds

The following figures show the spectral images of the shadows of the parallelepiped and the spheres that were located in space (Fig. 7).

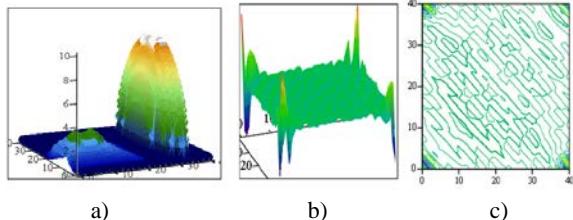


Fig. 7. Shadows of parallelepiped and spheres and the spectrum of their compatible shadows: a) shadows; b) three-dimensional image of the spectrum those shadows; c) a two-dimensional projection of the spectrum of shadows

Analysis of the spectra of hazardous and forbidden OC allows us to create an appropriate database for the further detection of OC of various shapes and complexity.

When using X-ray systems in order to provide automation of care and increase the reliability of decision-making on the presence of prohibited articles and substances in the OC, there are problems of identifying different forms and locations of the OC.

For this purpose, on the example of spectral detector model was constructed in the Matlab environment. In this case, the detection occurs regardless of the OC location and regardless of its shape and size.

The considered models are the shadows of two objects in a field with specified boundaries. One object is a regular square (this kind of can have a dynamite), and the other is a model of the machine gun (Fig. 8). Also, white Gaussian noise and a mixture of image and noise are modeled (Fig. 9). The developed program allows us to detect an OC with a given probability of false alarms for the corresponding threshold decision depending on the size of the OC. The program calculates the probability of correct detection of a signal from an OC.

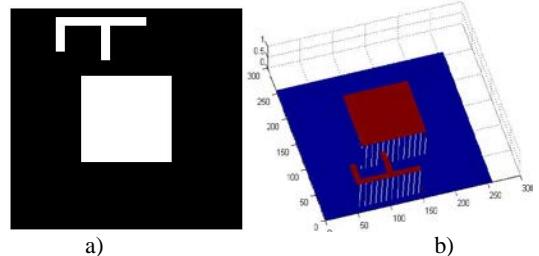


Fig. 8. Model of shadow OC

A mixture of useful signal and noise is shown in Fig. 9.

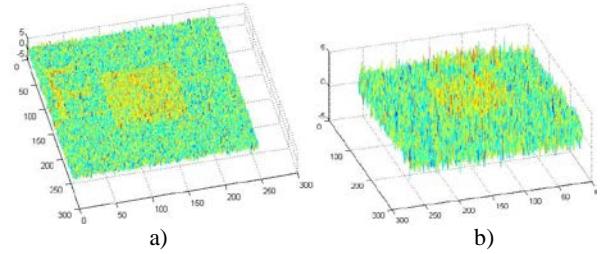


Fig. 9. A mixture of signal with noise in cases of signal-to-noise ratios equaled to 2 (a) and 0.5 (b)

The Neyman-Pearson criterion is applied for optimal detection of an OC. According to the Neyman-Pearson criterion, the threshold level  $V$  is determined from the condition that the probability of a correct detection  $D$  with the given probability of false alarm  $F$  was maximal. Hence, the optimal character of the Neyman-Pearson criterion is that it maximizes the probability of correct detection at a fixed probability of false alarms.

In addition, it should be noted that the program calculates the characteristics of the detection. An example of these characteristics is shown in Fig. 10.

On these graphs it is seen that when the decision threshold is reduced, the detection characteristic is more efficient, however, the probability of false detection is increased.

The analysis shows that the developed spectral detector has good detection characteristics even at low signal-to-noise ratios.

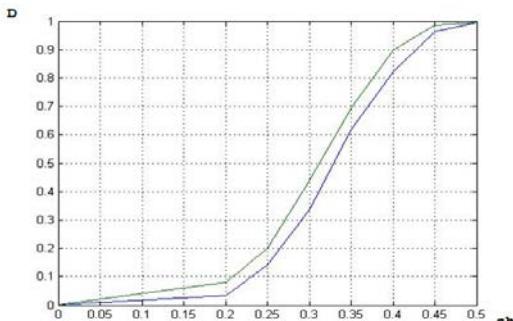


Fig. 10. Characteristics of signal detection for sample size 1000 and probabilities of false alarms  $F = 0.05$  and  $F = 0.03$

#### IV. CONCLUSION

The analysis of scientific publications has shown that the most effective methods for the detection and identification of hazardous OCs are transient multi-energy direct X-ray ones. They provide reliable detection of hazardous OCs. However, these methods are complicated, their implementation in the supervisory systems has a significant expenditure of material resources, and they do not work efficiently with dynamic OCs. At a high probability of correct detection to 0.99, there is a high probability of false alarms from 0.3 to 0.4.

The simulation shows that the simplest OC have shadows with transient characteristics, half-dooms, distortions of the type of the crater, where there are generally flat irradiating planes. Changing the irradiation angle changes the shadow to unrecognizability. To accurately identify the intended OC, it is necessary to automate the process of recognizing shadows, taking into account possible distances between the source, the OC and the screen-receiver, the irradiation angles, etc. Procedure for image processing consists of using given shape OC shadow to construct a two-dimensional spectrum and its subsequent use in developing the standard spectral detector. This detector is invariant with respect to the location of the OC in the working area. In order to solve the problem, a spectral detector model is developed using MatLab software environment. In this case, the detection occurs regardless of the OC location or its shape and size. It allows detecting dangerous objects with a high probability of correct detection and low probabilities of false positives (from 0.03 to 0.05).

#### REFERENCES

- [1] O. O. Semenov, *Theoretical foundations and principles of construction of technical devices for aviation security: training guide*, Kyiv, NAU, 2001, 214 p. (in Ukrainian).
- [2] V. N. Epifanov, et al. *Nondestructive inspection*, in 5 books. Book 1 *General questions. Control of Propagating Substances: Practicum*, Ed. by V.V. Sukhorukov, Moscow, Vysshaya shkola, 1993, 350 p. (in Russian).
- [3] *Nondestructive inspection*, in 5 books. Book 2. *Acoustic methods of control*, Ed. by V.V. Sukhorukov. Moscow, Vysshaya shkola, 1993, 380 p. (in Russian).
- [4] V. N. Epifanov, et al. *Nondestructive inspection*, in 5 books. Book 3. *Electrical, magnetic and eddy current testing methods and instruments: Practicum*, Ed. by V.V. Sukhorukov, Moscow, Vysshaya shkola, 1993, 420 p. (in Russian).
- [5] V. N. Epifanov, et al. *Nondestructive inspection*, in 5 books. Book 4. *Radiation control: Practicum*, Ed. by V.V. Sukhorukov, Moscow, Vysshaya shkola, 1992, 321 p. (in Russian).
- [6] V. V. Sukhorukov, et al. *Nondestructive inspection* in 5 books. Book 5 *Introspection and automation of control: Practicum*, Ed. by V.V. Sukhorukov, Moscow, Vysshaya shkola, 1993, 329 p. (in Russian).
- [7] *X-ray technology: Directory* in 2 books, Ed. by V. V. Klyuyev, Moscow, 1980, B.1, 431 p., B.2, 383 p. (in Russian).
- [8] S. Kolkoori, N. Wrobel, and U. Ewert, "A new X-ray backscatter technology for aviation security applications," in *2015 IEEE International Symposium on Technologies for Homeland Security (HST)*, Waltham, MA, USA, 14-16 April 2015, pp. 1-5.
- [9] S. Kolkoori, N. Wrobel, S. Hohendorf, and U. Ewert, "High Energy X-ray Imaging Technology for the Detection of Dangerous Materials in Air Freight Containers," in *2015 IEEE International Symposium on Technologies for Homeland Security (HST)*, Waltham, MA, USA, 14-16 April 2015, pp. 1-6.
- [10] S. Akcay, and T. P. Breckon, "An evaluation of region based object detection strategies within X-ray baggage security imagery," in *2017 IEEE International Conference on Image Processing (ICIP)*, Beijing, China, 17-20 Sept. 2017, pp. 1337-1341.
- [11] S. Akcay, M. E. Kundegorski, M. Devereux, and T. P. Breckon, "Transfer learning using convolutional neural networks for object classification within X-ray baggage security imagery," in *2016 IEEE International Conference on Image Processing (ICIP)*, Phoenix, AZ, USA, 25-28 Sept. 2016, pp. 1057-1061.
- [12] K. S. Kurachka, and I. M. Tsalka, "Vertebrae detection in X-ray images based on deep convolutional neural networks," in *2017 IEEE 14th International Scientific Conference on Informatics*, Poprad, Slovakia, 14-16 Nov. 2017, pp. 194-196.
- [13] Hewei Gao, Li Zhang, Zhiqiang Chen, Yuxiang Xing, Hui Xue, and Jianping Cheng, "Straight-Line-Trajectory-Based X-Ray Tomographic Imaging for Security Inspections: System Design, Image Reconstruction and Preliminary Results," *IEEE Transactions on Nuclear Science*, 2013, Volume 60, Issue 5, pp. 3955-3968.
- [14] W. Visser, et al. "Automated comparison of X-ray images for cargo scanning," in *2016 IEEE International Carnahan Conference on Security Technology (ICCST)*, Orlando, FL, USA, 24-27 Oct. 2016, pp. 1-8.
- [15] D. Mery, and A. K. Katsaggelos, "A Logarithmic X-ray Imaging Model for Baggage Inspection: Simulation and Object Detection," in *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Honolulu, HI, USA, 21-26 July 2017, pp. 251-259.
- [16] C. Meghare, and C. S. Gode, "Automated Detection Of Threat Object in X-ray Images of Baggage," *International Journal of Electrical, Electronics and Data Communication*, 2017, Volume 5, Issue 6, pp. 97-101.
- [17] A. A. Semenov, and L. Y. Tereshchenko, "Modeling of the visualization of the internal structure of objects of control," *Electronics and control systems*, 2008, № 1, pp. 144-148. (in Ukrainian).

# Simulation of Optimal Routes Passenger Transport

Oksana Bashutska<sup>1</sup>, Nadiia Panchuk<sup>2</sup>

1. Department of Economic Cybernetics and Informatics, Ternopil National Economic University, UKRAINE, Ternopil, 3 Peremoga square,  
email: [o.bashutska@gmail.com](mailto:o.bashutska@gmail.com)

2. Department of Economic Cybernetics and Informatics, Ternopil National Economic University, UKRAINE, Ternopil, 3 Peremoga square,  
email: [nadiapanchuk95@gmail.com](mailto:nadiapanchuk95@gmail.com)

**Abstract:** The purpose of the study is to create an economic and mathematical model for improving the performance of urban passenger transport and to make effective management decisions for planning and redevelopment of city routes. The research takes into account the current plans for the development of the city of Ternopil and the trends of urban passenger transportation.

**Keywords:** passenger transport, route, passenger transportation, optimal route, network, graph.

## I. INTRODUCTION

The main task of urban passenger transport is the provision of passenger transportation services. This topic is relevant as there is a need to create new routes and increase the competitiveness of existing urban passenger traffic through the provision of quality transport services. It is possible to improve their quality by improving technological and organizational transportation provision.

The theme of the study becomes particularly relevant, given the current plans for the development of the city of Ternopil and the trends of urban passenger transport. The optimal organization of urban passenger transport to date is not sufficiently explored. That is why we set ourselves the task of improving the organization of urban passenger transport, improving the quality of transport services for the city's residents, reducing transport tensions on the roads, improving the ecological situation and ensuring the economic efficiency of transport enterprises.

The following scientists worked on improving the efficiency of urban passenger transport: Afanasyev LL, Braylovsky MO, Butko T.V., Vorkut AI, Gavrilov EV, Geronimus B.L., Granovsky B. I., Dmitrichenko MF, Dolya V.K., Zabolotsky G.A., O.S. Ignatenko, V. V. Kobozev, A. Kotsyuk, P. Levkovets, P. Lopatin, L. P. Magnanti, T.L. Mun, V. V. Polischuk, A. Petrashevsky, Samoilov DS, Safronov Ye.A., Khabutdinov RA, MS Fiscells, Hasselstroem D., Mandl C., Nebelung H., Sonntag H. et al.

The city of Ternopil does not belong to large cities (more than 250 thousand people), nor to secondary (up to 250 thousand people), it is on the border between large and medium. For such a city, the time spent traveling from the place of residence to the place of work or training is 30-35 minutes, and for those who live in remote districts - 60-70 minutes. For this period in Ternopil 40% of passengers transport electric transport and 60% - bus. The total length of the contact network is more than 60 km, and the total length of trolleybus routes is approximately 150 km. The length of short trolleybus routes is 10-14 km, the average 15-19 km,

and long 20-23 km. The total length of bus routes is approximately 500 km. The average length of the bus route is 16.5 km.

We found that in the city of Ternopil, the longest bus route is the route number 18. In one direction, it travels about 15 km and it takes about 55 minutes. The longest trolleybus route is the route number 8. In one direction, it travels about 11 km and it takes about 52 minutes. In Ternopil, city passenger transportation provides 55 trolleybuses of large and especially large passenger capacity, as well as 210 buses of low passenger capacity, which are designed for 42 passengers.

The Trolley Park of Ternopil city contains 32 trolleybuses with a total passenger capacity of 100 passengers (TP14) and 23 trolleybuses with a passenger capacity of 150 passengers (TP15). However, in most of them the term of normative exploitation has expired and needs to be replaced. In order to fully update the trolleybus park, you need a lot of money, because new trolleybuses are very expensive (the new short Lviv or Lutsk trolleybus costs about 2 million UAH, and the big one - not less than 4 million UAH).

The bus fleet of the city consists of three types of buses: IVAN, Bogdan and Etalon. Most of these buses also need an update[2-5].

In Ternopil, motor transport complements electric transport due to duplication of routes. In peak hours, buses take up a significant number of passengers, and in the period when the number of passengers is small - contribute to reducing the number of trolleybuses on the line to save energy.

## II. THE METHOD OF OBTAINING OF THE OPTIMAL ROUTE

The terms of each routing task include a description of the network of communications, which determine the set of possible ways of following one or more moving objects. Typically, the structural parameters of the network remain unchanged from the beginning and until the end of the process of solving the problem.

The task of finding the optimal route in urban transport networks can only be solved by a complete overview of all possible options [7]. It is worth noting that the number of possible variants of route schemes is equal to  $2n \cdot (n - 1)$ , where  $n$  - the number of public transport stops. With increasing  $n$  this value is rapidly increasing, and already at  $n=10$  it is approximately  $1,24 \times 10^{27}$  variants.

It is clear that the complete overview of such an amount of options takes a lot of time and requires very powerful computing [6]. That is why we came up to solve this problem

from the expert point of view, that is, built the routes in order to cover as much territory as possible and at the same time minimize overlap with one transport route of another.

It is known that the base model for constructing an optimal transport network in the routing problem is a weighted graph  $H = (V, U)$  with a set of vertices  $V$  and a plurality of edges  $U$ . The vertex  $i \in V$ ,  $|V| = n$ , corresponds to city stops. The vertices  $i$  and  $j$  form in the graph  $H = (V, U)$  the edge  $\{i, j\}$  if they are represented by stops, directly connected segments of the road (trails), adjacent street crossings on a city map, etc[6].

Imagine an existing trolleybus route No. 9 (see Fig. 1) in the form of a graph. The vertices of this graph are the stops of the route, and the edges are the distance between the stops. Having analysed Figure 1, it can be argued that the current trolleybus route No. 9 has 29 stops, and its length is 21,795 km. Imagine an existing bus route No. 7 (see Fig. 2) in the form of a graph. The vertices of this graph are the stops of the route, and the edges are the distance between the stops. Having analysed Figure 2, it can be argued that the current trolleybus route No. 7 contains 25 stops, and its length is 6.95 km.

The calculation of the duration of the flight is carried out by the formula 1, which is presented below.

$$Tp = \frac{lm}{\beta \cdot Vt} + n_s \cdot t_s + t_{fs} \quad (1)$$

where  $lm$  – is the length of the route,

$\beta$  – runway coefficient (for Ternopil city  $\beta = 0,9$ ),

$Vt$  – technical velocity (average for Ternopil  $Vt = 25$  km/h),

$n_s$  – number of stops on the route,

$t_s$  – idle time at a stop (for the city of Ternopil  $t_s = 0,02$  h),

$t_{fs}$  – idle time at the final stop (for the city of Ternopil  $t_{fs} = 0,11$  h).

Calculate the duration of the voyage of the existing trolley route No. 9 by the formula 1:

$$Tp = \frac{21,795}{0,9 \cdot 25} + 29 \cdot 0,02 + 0,11 = 1,66 \text{ h.}$$

Calculate the duration of the voyage of the existing bus route number 7 by the formula 1:

$$Tp = \frac{6,95}{0,9 \cdot 25} + 29 \cdot 0,02 + 0,11 = 1 \text{ h.}$$

Thus, the duration of the voyage of the existing trolleybus route №9 is 1.66 hours, and the duration of the voyage of the existing bus route №7 is 1 hour.

The number of flights that is required for the smooth operation of urban passenger transport is calculated by the formula 2:

$$n_p = \frac{Tm}{Tp}, \quad (2)$$

where  $Tm$  – time on the route (for the city of Ternopil  $Tm = 17$  hours),

$Tp$  – the duration of the routes.

Calculate the optimal number of routes for the existing trolley route №9 for (2):

$$n_p = \frac{17}{1,66} = 10 \text{ routes}$$

Calculate the optimal number of routes for the existing bus route number 7 for (2):

$$n_p = \frac{17}{1} = 17 \text{ routes}$$

Thus, the required number of routs for the existing trolleybus route number 9 is 10, and the required number of routs for the existing bus route number 7 is 17.

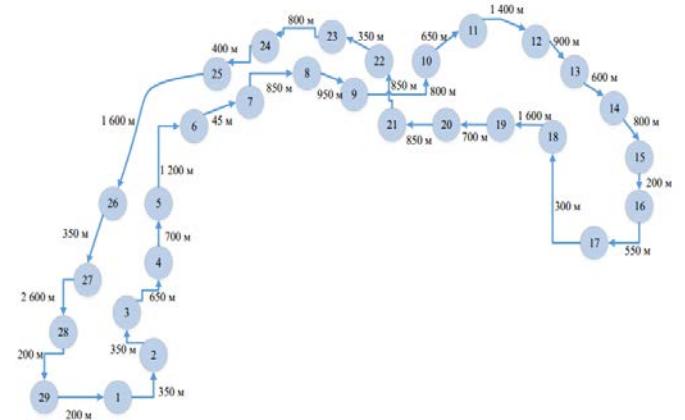


Fig. 1 Active trolleybus route №9

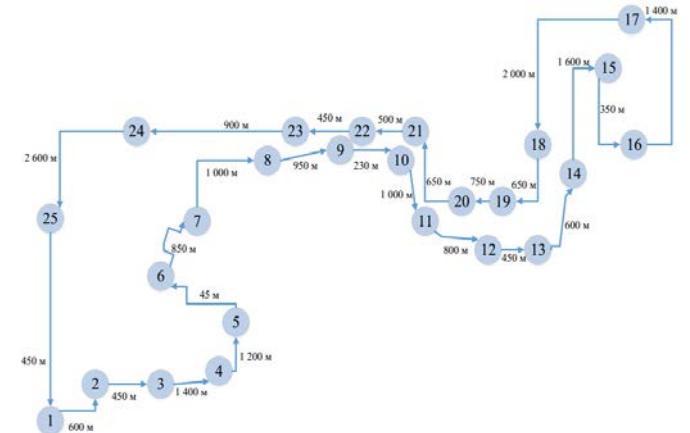


Fig. 2 Active bus route №7

The need to build new optimal routes in the city of Ternopil is due to the fact that a new micro district is being built, which will provide the city with a large passenger traffic (approximately 43 284 people/month). This is due to the fact that in the "Warsaw neighbourhood" will be built: a large sports complex and swimming pool, shopping and entertainment centre, a new bus station, multi-level parking, houses, school and kindergarten.

It should be emphasized that for the laying of routes in the area, it is necessary first to lay 750 m road and 1850 m contact network. It will bring the city's expenses in the amount of 6008000 UAH. However, these costs will quickly pay off.

Imagine a new trolleybus route №9A (see Figure 3) in the form of a graph. Having analyzed Figure 3, it can be argued that the new trolleybus route №9A contains 29 stops, and its length is 20,755 km.

Imagine a new bus number 7 (see Figure 4) in the form of a graph. After analyzing Figure 4, it can be argued that the new bus №7 has 27 stops, and its length is 20,435 km.

Calculate the duration of the route of the new trolley route №9A for (1):

$$Tp = \frac{20,755}{0,9 \cdot 25} + 29 \cdot 0,02 + 0,11 = 1,62h.$$

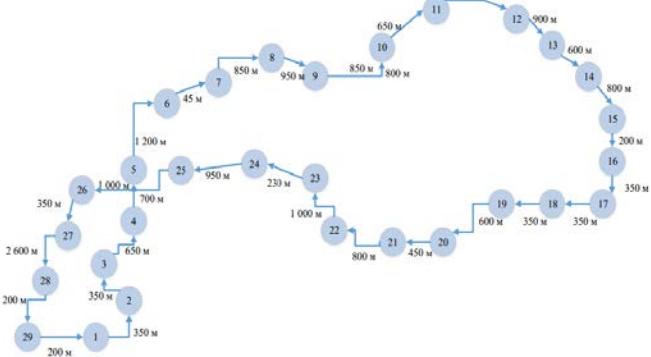


Fig. 3 New optimized trolleybus route №9A

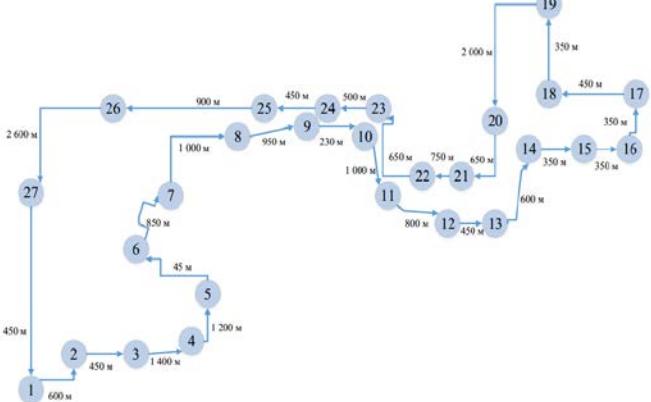


Fig. 4 New optimized bus route №7

Calculate the duration of the routes of the new bus route number 7 for (1):

$$Tp = \frac{20,435}{0,9 \cdot 25} + 29 \cdot 0,02 + 0,11 = 1,6 h.$$

Thus, the duration of the route of the new trolley route №9A is 1.66 hours, and the duration of the flight of the new bus route №7 is 1 hour.

Calculate the optimal number of routes for the new trolley route №9A for (2):

$$n_p = \frac{17}{1,62} = 11 \text{ routes}$$

Calculate the optimal number of routes for the new bus route №7 for (2):

$$n_p = \frac{17}{1,6} = 11 \text{ routes}$$

Thus, the required number of routes for the new trolley route №9A and the new bus route №7 is 11.

Determine the planned volume of passenger transportation per day:

$$Q_{daily} = \frac{Q_p}{D_c}, \quad (3)$$

where  $Q_p$  – is the volume of passengers transported per year;

$D_c$  – calendar number of days in a year.

Planned volume of passenger traffic per day with new optimal routes:

$$Q_{daily} = \frac{43\,284 \cdot 12}{365} = 1\,423 \text{ pass.}$$

Determine the daily passenger traffic by the formula:

$$P_{daily} = Q_{daily} \cdot l_{med}, \quad (4)$$

where  $l_{med}$  – is the average distance over which the passenger overcomes (for the city of Ternopil  $l_{med} = 2 \text{ km}$ ).

$$P_{daily} = 1\,423 \cdot 2 = 2\,846 \text{ pass.km}$$

The coefficient of variability of passengers is determined by the formula:

$$\eta_v = \frac{lm}{l_{med}} \quad (5)$$

Determine the coefficient of passenger variation for the new trolley route No. 9A:

$$\eta_v = \frac{20,755}{2} = 10,38$$

Determine the coefficient of variation of passengers for the new bus number 7:

$$\eta_v = \frac{20,435}{2} = 10,22$$

Determine the maximum daily productivity of vehicles:

$$W_Q^{max} = q \cdot Y_{max} \cdot n_p \cdot \eta_v,$$

where  $q$  is the nominal capacity of the vehicle (the "Etalon" bus is 42; the trolleybus "14Tr" is 100; the trolleybus "15Tr" is 172);

$Y_{max}$  - coefficient of passenger capacity utilization (for the city of Ternopil  $Y_{max} = 0,5$ ).

Since, in Ternopil there are two types of trolleybuses with different passenger capacity, we will determine the maximum daily productivity for each of them with the help of (6).

For trolleybus "14Tr" of the new trolley route №9A:

$$W_Q^{max} = 100 \cdot 0,5 \cdot 11 \cdot 10,38 = 5\,709 \text{ pass.}$$

For trolleybus "15Tr" of the new trolley route №9A:

$$W_Q^{max} = 172 \cdot 0,5 \cdot 11 \cdot 10,38 = 9\,820 \text{ pass.}$$

For the bus "Etalon" of the new bus route number 7:

$$W_Q^{max} = 42 \cdot 0,5 \cdot 11 \cdot 10,22 = 2\,361 \text{ pass.}$$

Determine the required number of vehicles:

$$A_{pc} = \frac{Q_{daily}}{W_Q^{max}} \quad (7)$$

For trolleybus "14Tr" of the new trolley route №9A:

$$A_{pc} = \frac{1\,423}{5\,709} \approx 1 \text{ trolleybus}$$

For trolleybus "15Tr" of the new trolley route №9A:

$$A_{pc} = \frac{1\,423}{9\,820} \approx 1 \text{ trolleybus}$$

For the bus "Etalon" of the new bus route number 7:

$$A_{pc} = \frac{1\,423}{2\,361} \approx 1 \text{ bus}$$

So, for optimal operation of the transport in the "Warsaw district" it is expedient to launch 1 bus "Etalon", 1 trolley bus "14Tr" and 1 trolley "15Tr".

We will calculate what revenue will get vehicles per day for this multiply the maximum performance of the vehicle by the fare in it:

For trolleybus "14Tr" of the new trolley route №9A:

$$\text{Revenue} = 5\,709 \cdot 3 = \text{UAH } 17\,127.$$

For trolleybus "15Tr" of the new trolley route №9A:

$$\text{Revenue} = 9\,820 \cdot 3 = \text{UAH } 29\,460$$

For the bus "Etalon" of the new bus route number 7:

$$\text{Revenue} = 2\,361 \cdot 4 = \text{UAH } 9\,444$$

The total revenue that vehicles will receive per day is UAH

56030. However, taking into account the privileged travel, it will amount to UAH 47626.

Develop a visualization of the transport system and interactive surveillance, which will allow you to see the traffic of vehicles on the route in real time from stop to stop.

This interactive model will, unlike the existing ones, automatically take into account the location of the vehicle on the route, and will also provide instant detection of deviations from normal traffic.

The interactive transport network, constructed taking into account the adjusted routes for the new "Warsaw" district in the city of Ternopil, is presented in Figures 5 and 6.

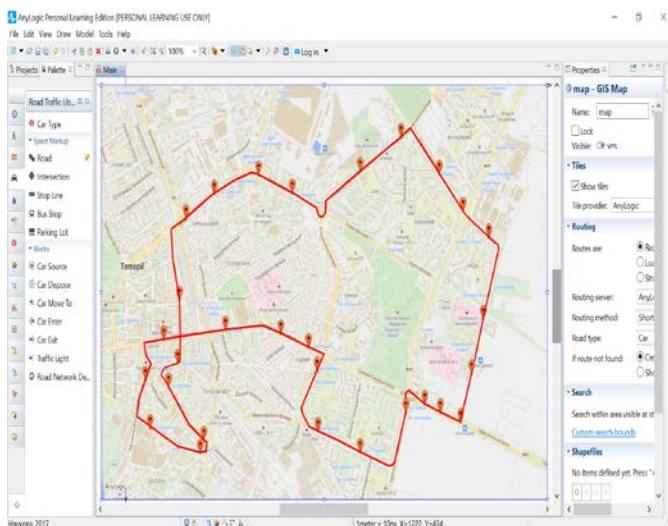


Fig. 5 Interactive transport network of optimized trolleybus route №9

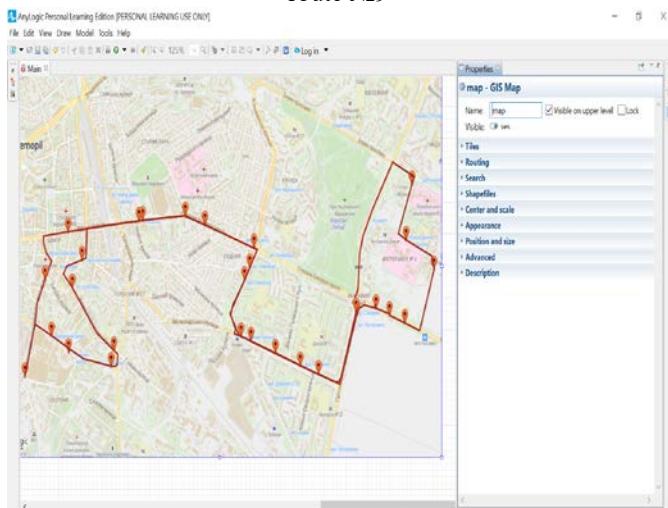


Fig. 6. Interactive transport network of optimized bus route №7

The constructed model of the interactive transport network can be used as a tool for developing new and improving existing urban passenger transport routes. It allows you to predict the number of vehicles on the route, the length of routes, the intervals between vehicles. It also allows you to take into account speed mode and change of traffic rules (for example, the appearance of new road signs on routes).

### III. CONCLUSION

As a result of the research carried out, the visualization of the transport system and interactive surveillance, which allows you to see the movement of vehicles on the route in real time from stop to stop.

This interactive model will, unlike the existing ones, automatically take into account the location of the vehicle on the route, and will also provide instant detection of deviations from normal traffic.

An interactive transport network, built on tailored routes for a new micro district in the city of Ternopil.

The constructed model of an interactive transport network can be used as a tool for developing new and improving existing urban passenger transport routes. It allows you to predict the number of vehicles on the route, the length of routes, the intervals between vehicles. It also allows you to take into account speed mode and change of traffic rules (for example, the appearance of new road signs on routes).

### REFERENCES

- [1] Grigorkiv V. S. Economic cybernetics: Teaching. manual - Chernivtsi: Ruta, 2006. - 9 p.
- [2] Classification and system properties [Electronic resource]: - Access mode: <http://buklib.net/books/22350/>
- [3] Law of Ukraine "On Transport" [Electronic resource]: - Mode of access: <http://zakon3.rada.gov.ua/laws/show/232/94-%D0%B2%D1%80>
- [4]. Internet article "Trolley-bus fleet of Ternopil plan to update in 2 years" [Electronic resource]: - Access mode: <http://www.0352.ua/y/article / 277181>
- [5] News of the official site of the Municipal Enterprise "Ternopillectrotrans" [Electronic resource]: - Access mode: <http://elektrotrans.te.ua/routes/>
- [6] T.L. Magnanti, R.T. Wong, "Network design and transpotation planning: models and algorithms", *Transportation Science*. - 1984. - No. 18 (1), pp. 1-55.
- [7] B.L. Gerominus and L.V. "Chaparin Economic-mathematical methods in planning on motor transport" Moscow: Transport, 1988. - 192 p. (in Russian)

# Inductive Modeling of Amylolytic Microorganisms Quantity in Copper Polluted Soils

Olha Moroz, Volodymyr Stepashko

Department for Information Technologies of Inductive Modelling  
 International Research and Training Centre for Information Technologies and Systems of the NASU and MESU  
 Glushkov Avenue 40, Kyiv, 03680, Ukraine  
 olhahryhmoroz@gmail.com, stepashko@irtc.org.ua

**Abstract:** The article presents the application results of the combinatorial-genetic algorithm COMBI-GA for modeling from experimental data of amylolytic microorganism quantity in a soil plot contaminated by copper. The constructed nonlinear mathematical models describe dependence of microorganisms concentration in soil from basic vital environmental factors and the copper concentration in the soil.

**Keywords:** GMDH, combinatorial algorithm COMBI, genetic algorithm GA, hybrid algorithm COMBI-GA, amylolytic microorganisms, soil, heavy metals, inductive modelling.

## I. INTRODUCTION

One of the main components of most ecosystems is soil, in which microorganisms play an important role in the evolution and formation of fertility. Anthropogenic pollution of the biosphere affects all living components of biogeocenoses, including soil microorganisms [1].

Modern agroecosystems are subject to the considerable technogenic influence resulting frequently in pollution of arable soils. Pollutants influence negatively on soil microbiotics, which causes the necessity to carry out long-term observations after its current status. In parallel with monitoring, there is the task of determination of critical deviations and prediction of the microbiotum state dependently on pollutants concentration in soil. Solving this task is possible only by formalization of monitoring data in the form of mathematical models [2].

Investigation of the effect of specific anthropogenic factors, such as heavy metals, on microbial community functioning is very important. Negative influence of heavy metals on microbial kenosis, soil and biological activity is well-known. In this connection, the organization and carrying out of regular control of soil state for the purpose of critical situation detecting and forecasting are actual.

Thus the dynamics of change of microorganism quantity in soil is studied under influence of different ecological factors. Modeling from observation data is the necessary condition of the ecological monitoring as it allows operative estimating current ecological situations and forecasting their development.

In this research we find models of dependence of microorganisms functioning from supporting weather

conditions, this means kind of models, where inputs are hydrometeorological variables and outputs ecological indexes. Such models can be used both for restoring the omitted data and for forecasting the development of the controlled ecological processes on the basis of weather terms prediction. For construction of such models we use the Group Method of Data Handling (GMDH) as the most effective method for the analysis, modeling and forecasting of complex processes from experimental data under conditions of incompleteness of a priori information and short samples given.

To analyse of soil microbiotics of dark gray podzol soil (Kyiv region) polluted with heavy metals, the automated system of simulation ASTRID [3] with different algorithms of GMDH was initially used.

In [4] the application results of the hybrid combinatorial-genetic algorithm COMBI-GA [5, 6] for finding optimal linear models on the basis of observations of the change in the number of amylolytic microorganisms in the soil contaminated by copper are presented. But to solve prediction tasks we need more accurate approximation. So, in this paper new results of the application of the COMBI-GA algorithm for building optimal nonlinear models are presented.

Section II of this paper describes briefly task of modelling. Section III considers hybrid combinatorial-genetic GMDH algorithm COMBI-GA and their features. Section IV presents modelling results.

## II. TASK OF MODELING

Experiments regarding functioning amylolytic microorganisms under copper contamination were carried out on small plots in deep-gray podzolic soil (Kyiv region). The traditional chart [7] of experiments was used: several plots of the same soil type were selected for experiments and a plot remained as control non-contaminated one.

Model contamination of soil was carried out by the annual one time bringing in soil solutions of  $Cu^{2+}$  salts at the beginning of a vegetation season. The amount of the applied metal (computed as content of their ions) corresponds to contamination doses of 2 maximum permissible concentrations (MPC). The soil pieces for the analyses was taken during vegetation periods from 1993 to 1996 from the arable layer depth (0-20 cm) approximately in 2nd, 30th and 90th day after bringing of the metal salt. It was hence

received three measuring points during four years or 12 points together.

The amount of amylolytic microorganisms in the control and polluted soils was determined by the method of sowing of soil suspension on a nourishing medium consisting of a starch-ammonia agar.

Based on observations data of below listed variables, the linear mathematical models were built in [8] for description of the amylolytic bacteria quantity changes in control and polluted by the heavy metal soils.

As input (independent) variables (factors) for construction of models were used: concentration of mobile forms of  $Cu^{2+}$ , decade average values of temperature, humidity of soil and air, and number of microorganisms in soil of control unpolluted plot. Quantity data of amylolytic microorganisms were output (dependent) variables in plots with model pollution of soil by copper salt. Based on the obtained data, the models of microorganisms number in soil were built.

For construction of model of changing quantity of amylolytic microorganisms such list of input variables was formed:  $x_1$  – quantity of microorganisms in the control plot (millions in 1 g of dry soil);  $x_2$  – concentration of copper (mg/cg soil);  $x_3$  – number of days from the date of pollution;  $x_4$  – current decade average temperature of air (oC);  $x_5$  – previous decade average air temperature (oC);  $x_6$  – current decade average humidity of air (%);  $x_7$  – previous decade average humidity of air (%)

A definition of the inductive modelling problem in this task may be done as follows. Let us given: a data set of  $n$  observations after 7 inputs  $x_1, x_2, \dots, x_7$  and one output  $y$  variables. The GMDH task is to find a model  $y=f(x_1, x_2, \dots, x_7, \theta)$  with minimum value of a given model quality criterion  $C(f)$ , where  $\theta$  is unknown vector of model parameters. The optimal model is defined as  $f^*=argmin_{\Phi} C(f)$ , where  $\Phi$  is a set of models of various complexity,  $f \in \Phi$ .

### III. HYBRIDS GMDH-GA ALGORITHM

The genetic algorithm [9] is one of the meta-heuristic procedures of global optimization constructed as a result of generalization and simulation in artificial systems of such properties of living nature as natural selection, adaptability to changing environmental conditions, inheritance by offspring of vital properties from parents.

Since GA is based on the principles of biological evolution and genetics, biological terms are used actively (and sometimes incorrectly) to describe them. Here are some of these terms. *Individual* is the potential solution to the problem; *population* is a set of individuals; *offspring* is usually improved copy potential solution (father); *fitness* is usually a quality characteristic of the solution. *Chromosome* is encoded data structure of an individual in the form of an array of fixed lengths. In the simplest case it's a binary string of fixed length. The *gene* is an element of this array.

Formally, GA can be represented in such a way:

$$GA = \{P_0, M, L, F, G, s\},$$

where  $P_0 = (a_1^0, \dots, a_M^0)$  is an initial population;  $a_i^0$  is an individual of this population treated as a candidate for the solution of the optimization problem presented in the form of

a chromosome;  $M$  is the population size (integer number);  $L$  is the length of each chromosome of the population (integer number);  $F$  is a fitness function of an individual;  $G$  is a set of genetic operators;  $s$  is the algorithm stopping rule.

As input data for any GA initial population  $P_0$ , a finite set of chromosomes is used each of which represents a potential solution of the problem. Then the first population of offspring  $P_1$  is formed from the parent chromosomes  $P_0$  using some genetic operators, similarly the next population  $P_2$  is formed from the population  $P_1$  and so on. The process continues until the specified stopping rule of the algorithm will be satisfied.

An important feature of the GA work is that with each step the mean FF value of the current population improved and converges to the solution of the optimization problem.

The effectiveness of GA's work depends on the method of encoding genes, the composition of the initial population used by genetic operators, GA parameters, such as population size, number of chromosomes selected during selection and for crossover, probability of using genetic operators. The most important in GA are genetic operators especially the selection of which stores a certain amount of chromosomes with the best values of FF for each iteration of GA, and the operators of the creation of new offspring-chromosomes such as crossover and mutation. The crossover operator creates offspring by exchanging genetic material between the parent chromosomes, and the mutation operators by changing one chromosome in accordance with certain rules.

Formally, the hybrid of COMBI [10] and GA algorithm can be defined as follows:

$$COMBI-GA = \langle Z, y, f, X, D, CR, P_0, H, M, G, k, F \rangle,$$

where  $Z[n \times r]$  is the measurement matrix of input variables of an object,  $r$  and  $n$  are numbers of inputs and measurements respectively;  $y[n \times 1]$  is vector of measurements of an output variable;  $f[m \times 1]$  is vector of a given  $m$  base functions of input variables;  $X[n \times m]$  is the measurement matrix of base set of arguments;  $D$  is a given rule of dividing matrix  $X[n \times m]$  and vector  $y[n \times 1]$  to testing  $A$  and checking  $B$  parts;  $CR$  is an external selection criterion (as fitness function) based on dividing the sample ( $X, y$ );  $P_0$  is a set of model structures of GA initial population consisting of binary chromosomes (encoded structure of partial models);  $H$  is size of initial population of models,  $H < m$ ;  $M$  is size of any next population,  $M > H$ ;  $G$  is set of genetic operators;  $k$  is stopping rule of GA;  $F$  is number of best partial models (freedom of choice) monitored during all iterations of the algorithm,  $1 < F \leq H$ .

This algorithm consists of the following steps:

*Step 1.* Calculating the matrix of the base set of arguments  $X[n \times m]$  using the input matrix  $Z$  and the vector of base functions  $f$  and dividing it and the output vector of measurements  $y[n \times 1]$  according to the rule  $D$  in testing  $X_A[n_A \times m]$  and checking  $X_B[n_B \times m]$  submatrices ( $n_A + n_B = n$ ). Obviously, in the case of linear polynomial, matrices  $X$  and  $Z$  are identical ( $m = r$ ).

*Step 2.* Random generating the initial population  $P_0$  of the genetic algorithm.

*Step 3.* Calculating the coefficients of each partial model by LSM or another method using the training matrix of base

arguments  $X_A$  and output vector  $y_A$ .

*Step 4.* Calculating the value of an external criterion  $CR$  (as the GA fitness function) for each partial model using the checking matrix  $X_B$  and output  $y_B$ .

*Step 5.* Forming the current population of partial models (chromosomes) of the size  $H$  with better criterion values to form the next offspring. In addition, selection the best  $F$  partial models that are potential solutions of the task of model building.

*Step 6.* Forming new population of  $M$  individuals applying genetic operators of crossover and mutation to individuals of the current population.

*Step 7.* Checking a given GA stopping rule. If it is satisfied, then go to step 8, otherwise go to step 3.

*Step 8.* Choosing  $F$  best models from the current population of the size  $H$ .

*Step 9.* The end.

#### IV. MODELING RESULTS

Based on experimental data, models of quantity of amylolytic microorganisms were built in the control as well as in the copper polluted soils. In all cases we use the COMBI-GA algorithm with the following division of all data sample (12 points of observation during 4 years for the vegetation period 1993-1996): 6 points (2 years) as training set  $A$ , 3 points (1 year) as checking set  $B$ , and 3 points (1 year) as validation set  $C$ .

*Results in the class of linear models.* In case of linear modelling, the quantity of amylolytic in a control soil  $Y_{contr} = x_1$  is described by the model obtained in [4]:

$$Y_{contr} = 0,2136x_4 - 0,7149x_5 + 0,5412x_7.$$

Models for the quantity of microorganisms in polluted soil were built taking into account the quantity of amylolytic microorganisms at the observation of the control soil  $x_1$ .

The linear model for quantity of amylolytic microorganisms in the copper polluted soil [4]:

$$Y = 0,743x_1 + 1,6516x_2 - 0,9182x_5 - 0,2845x_7$$

Proper graphs of experimental and model data are given on Fig. 2. The characteristics of accuracy of these models are presented in the table below. This accuracy level is insufficient for quality monitoring needs. That is why we decide to build more complex nonlinear (polynomial) models. The results of this stage of modeling are presented below.

*Results in the class of nonlinear models.* In the case of nonlinear modelling, the quantity of amylolytics in a control soil is described by the model:

$$x_1 = -0,19297x_5 - 2,0976x_7 + 0,130x_4x_5 + 0,111x_5x_7$$

Fig. 3 shows graphs of measured and modeled data .

The model of dynamics of changing the amylolytic microorganisms in soil polluted by copper:

$$Y = 1,1759x_5 + 0,2159x_7 + 0,1638x_1x_5 - 0,0261x_2x_7 - 0,1252x_4x_5$$

The characteristics of all obtained models quality are presented in the table calculated according to next formulas:

$$MSE = \sqrt{\frac{1}{12} \sum_{i=1}^{12} (x_i - \bar{x})^2}, AR_B = \|y_B - X_B \hat{\theta}_A\|^2.$$

The designation "Valid. err." means error on the independent validation set  $C$  calculated like  $AR_B$ .

|             | Linear case |           | Nonlinear case |           |
|-------------|-------------|-----------|----------------|-----------|
|             | Measured    | Predicted | Measured       | Predicted |
| MSE         | 0,738       | 0,814     | 0,124          | 0,138     |
| AR          | 0,138       | 0,214     | 0,058          | 0,061     |
| Valid. err. | 0,120       | 0,157     | 0,094          | 0,110     |

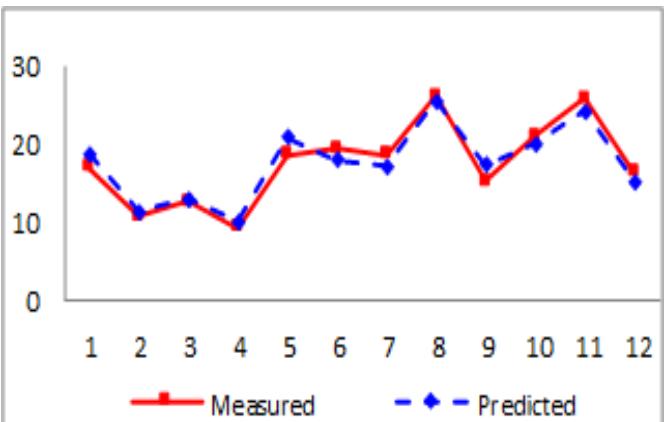


Fig. 1 Graphs of quantity change of amylolytic microorganisms on the control plot (linear model).

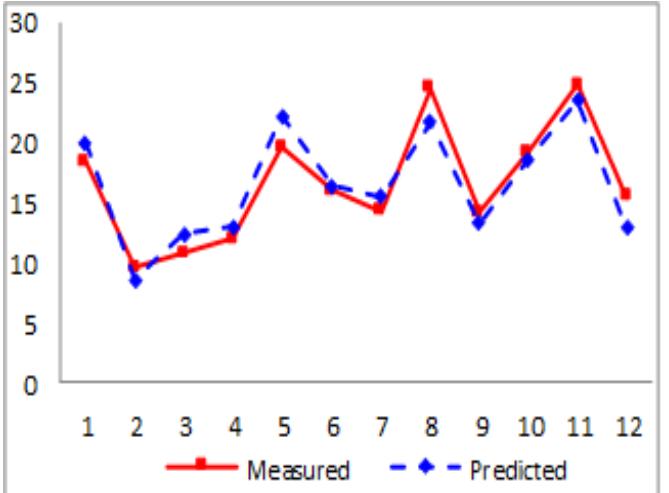


Fig. 2 Graphs of quantity change of amylolytic microorganisms on the plot polluted by copper (linear model).

Proper graphs of experimental and model data are given on Fig. 4. These graphs shows that in most points the data measured and predicted by the model coincide, that is the models adequately represents the change of microorganisms quantity. Three last three validation points on the graphs testify good results of models verification in the forecasting mode. Some distinctions can be accounted for by spatial heterogeneity of soil and other terms what could cause irregular variability of quantity.

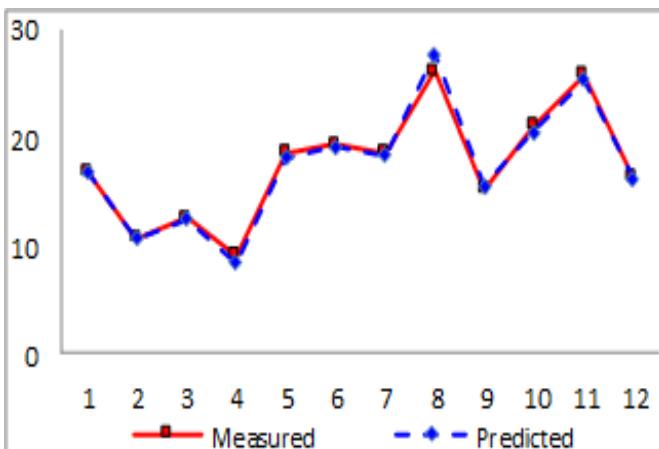


Fig. 3 Graphs of change of quantity of amylolytic microorganisms on the control plot (nonlinear model)

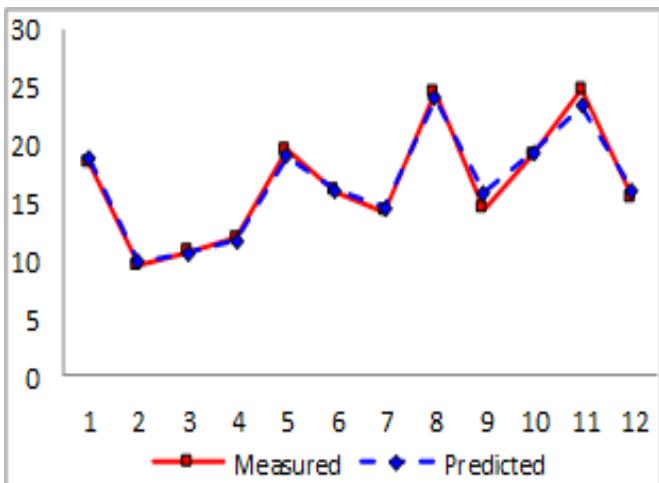


Fig. 4 Graphs of change of quantity of amylolytic microorganisms on the plot polluted by copper (nonlinear model)

As it is evident from the results, the functioning of amylolytic bacteria in soil is substantially influenced by the temperature and humidity of air.

In nonlinear case we obtain much more accurate results that can further help to effectively solve different ecological tasks based on microbial monitoring.

## V. CONCLUSION

The carried out research manifests the possibility of formalization of the given ecological observations by construction of mathematical models. For the modeling of quantity of microorganisms in soil, application of inductive approach on the basis of GMDH is effective.

The obtained nonlinear models in a high degree coincide with experimental data that enables to use them in the system of the experiments for estimation of degree of soil contamination, renewal of intermediate or omitted data and operative forecasting the dynamics of microorganisms under various ecological conditions. Equally, these models will be helpful also for the data restoration with the purpose of obtaining the uniform series of ecological observations.

## REFERENCES

- [1] K.I. Andreyuk, H.O. Iutynska, A.F. Antypchuk et al, "The functioning of soil microbial communities under conditions of anthropogenic load," K.: *Oberehy*, 2001, 240 p.
- [2] H. G. Schlegel, "General microbiology," 7th edition, Cambridge University Press, 1993, 655 p.
- [3] V.S. Stepashko, Yu.V. Koppa, "Experience of the ASTRID system application for the modeling of economic processes from statistical data," *Cybernetics and computing technique*, vol. 117, pp.24-31, 1998. (in Russian)
- [4] G. Iutynska, O.Moroz, "Inductive modeling of changes of amylolytic microorganisms on polished surface tuber," Inductive modeling of complex systems, *IRTC ITS NASU*, Kyiv, 2017, vol. 9, pp. 85-91.
- [5] O. Moroz, V. Stepashko, "Hybrid Sorting-Out Algorithm COMBI-GA with Evolutionary Growth of Model Complexity," *Advances in Intelligent Systems and Computing II* / N. Shakhovska, V. Stepashko, Editors, AISC book series, Berlin: Springer Verlag, vol. 689, pp. 346-360, 2017.
- [6] O.H. Moroz, "Sorting-Out GMDH algorithm with genetic search of optimal mode," *Control Systems and Machines*, no. 6, pp. 73-79, 2016. (In Russian)
- [7] E.I. Andreyuk, G.A. Iutynska, Z.V. Petrusha, "Homeostasis of microbial of soils polluted by heavy metals," *Mikrobiol. Journ*, vol 61, no. 6, pp.15-21, 1991. (in Russian)
- [8] A.G. Ivakhnenko, V.S. Stepashko, "Noise-Immunity of Modeling," Kiev: *Naukova Dumka*, 1985, 216 p. (In Russian)
- [9] J. Holland, "Adaptation in natural and artificial systems, An introductory analysis with application to biology, control, and artificial intelligence," *University of Michigan, Computers*, 1975, 183 p.
- [10] V.S. Stepashko, "Combinatorial Algorithm of the Group Method of Data Handling with Optimal Model Scanning Scheme," *Soviet Automatic Control*, vol 14, no 3, pp. 24-28, 1981.

# Generalized Transport Equation with Fractality of Space-Time. Zubarev's NSO Method

Petro Kostrobij<sup>1</sup>, Bogdan Markovych<sup>1</sup>, Olexandra Viznovych<sup>1</sup>, Mykhailo Tokarchuk<sup>1,2</sup>

1. Lviv Polytechnic National University, 12 S. Bandera str., 79013, Lviv, Ukraine, email: [bohdan.m.markovych@lpnu.ua](mailto:bohdan.m.markovych@lpnu.ua)

2. Institute for Condensed Matter Physics of the National Academy of Sciences of Ukraine, 1 Svientsitskii str., 79011, Lviv, Ukraine

**Abstract:** We presented a general approach for obtaining the generalized transport equations with fractional derivatives by using the Liouville equation with fractional derivatives for a system of classical particles and the Zubarev nonequilibrium statistical operator (NSO) method within Renyi statistics. Generalized Cattaneo-type diffusion equations with taking into account fractality of space-time are obtained.

**Keywords:** fractional derivative, diffusion equation.

## I. INTRODUCTION

The fractional derivatives and integrals [1]–[4] are widely used to study anomalous diffusion in porous media, in disordered systems, in plasma physics, in turbulent, kinetic, and reaction-diffusion processes, etc. [5], [6]. In Ref. [5], [6], we discussed various approaches to obtaining the transport equations with fractional derivatives. It is important to note that, for the first time, in Refs. [7]–[10], Nigmatullin received diffusion equation with the fractional time derivatives for the mean spin density [7], the mean polarization [8], and the charge carrier concentration [9]. In Ref. [10], justification of equations with fractional derivatives is given, and the time irreversible Liouville equation with the fractional time derivative is provided. In our recent work [5], by using NSO method [11], [12] and the maximum entropy principle for the Renyi entropy, we obtained the generalized (non-Markovian) diffusion equation with fractional derivatives. The use of the Liouville equation with fractional derivatives proposed by Tarasov in Refs. [13], [14] is an important and fundamental step for obtaining this equation. By using NSO method and the maximum entropy principle for the Renyi entropy, we found a solution of the Liouville equation with fractional derivatives at a selected set of observed variables. We chose nonequilibrium average values of particle density as a parameter of reduced description, and then we received the generalized (non-Markovian) diffusion equation with fractional derivatives. In the next section by using Ref. [5], new non-Markovian diffusion equations for particles in a spatially heterogeneous environment with fractal structure are obtained. Different models of frequency-dependent memory functions are considered, and the diffusion equations with fractality of space-time are obtained.

## II. LIOUVILLE EQUATION WITH FRACTIONAL DERIVATIVES FOR SYSTEM OF CLASSICAL PARTICLES

We use the Liouville equation with fractional derivatives obtained by Tarasov in Refs. [14] for a nonequilibrium particle function  $\rho(x^N; t)$  of a classical system

$$\frac{\partial}{\partial t} \rho(x^N; t) + \sum_{j=1}^N D_{\vec{r}_j}^\alpha (\rho(x^N; t) \vec{v}_j) + \sum_{j=1}^N D_{\vec{p}_j}^\alpha (\rho(x^N; t) \vec{F}_j) = 0, \quad (1)$$

where  $x^N = x_1, \dots, x_N$ ,  $x_j = \{\vec{r}_j, \vec{p}_j\}$  are dimensionless generalized coordinates,  $\vec{r}_j = (r_{j1}, \dots, r_{jn})$ , and generalized momentum,  $\vec{p}_j = (p_{j1}, \dots, p_{jn})$ , [14] of  $j$  th particle in the phase space with a fractional differential volume element [13], [15]  $d^\alpha V = d^\alpha x_1 \dots d^\alpha x_N$ . Here,  $m = Mr_0/(p_0 t_0)$ ,  $M$  is the mass of particle,  $r_0$  is a characteristic scale in the configuration space,  $p_0$  is a characteristic momentum, and  $t_0$  is a characteristic time,  $d^\alpha$  is a fractional differential [15],

$$d^\alpha f(x) = \sum_{j=1}^{2N} D_{x_j}^\alpha f(x) (dx_j)^\alpha,$$

where

$$D_x^\alpha f(x) = \frac{1}{\Gamma(n-\alpha)} \int_0^x \frac{f^{(n)}(z)}{(x-z)^{\alpha+1-n}} dz \quad (2)$$

is the Caputo fractional derivative, [1], [2], [16], [17]  $n-1 < \alpha < n$ ,  $f^{(n)}(z) = d^n f(z)/dz^n$  with the properties  $D_{x_j}^\alpha 1 = 0$  and  $D_{x_j}^\alpha x_l = 0$ , ( $j \neq l$ ).  $\vec{v}_j$  are the fields of velocity,  $\vec{F}_j$  is the force field acting on  $j$  th particle. If  $\vec{F}_j$  does not depend on  $\vec{p}_j$ ,  $\vec{v}_j$  does not depend on  $\vec{r}_j$ , and the Helmholtz conditions, we get the Liouville equation in the form

$$\frac{\partial}{\partial t} \rho(x^N; t) + iL_\alpha \rho(x^N; t) = 0, \quad (3)$$

where  $iL_\alpha$  is the Liouville operator with the fractional derivatives,

$$iL_\alpha \rho(x^N; t) = \sum_{j=1}^N [D_{\vec{p}_j}^\alpha H(\vec{r}, \vec{p}) D_{\vec{r}_j}^\alpha - D_{\vec{r}_j}^\alpha H(\vec{r}, \vec{p}) D_{\vec{p}_j}^\alpha] \rho(x^N; t). \quad (4)$$

where  $H(\vec{r}, \vec{p})$  is a Hamiltonian of a system with fractional derivatives [13]. A solution of the Liouville equation (3) will be found with Zubarev's NSO method [11]. After choosing parameters of the reduced description, taking into account projections we present the nonequilibrium particle function  $\rho(x^N; t)$  (as a solution of the Liouville equation) in the general form

$$\begin{aligned} \rho(x^N; t) &= \rho_{rel}(x^N; t) \\ &- \int_{-\infty}^t e^{\epsilon(t'-t)} T(t, t') (1 - P_{rel}(t')) iL_\alpha \rho_{rel}(x^N; t') dt', \end{aligned} \quad (5)$$

where  $T(t, t') = \exp_+ \left[ - \int_{t'}^t (1 - P_{\text{rel}}(t')) iL_\alpha dt' \right]$  is the evolution operator in time containing the projection,  $\exp_+$  is ordered exponential,  $\varepsilon \rightarrow +0$  after taking the thermodynamic limit,  $P_{\text{rel}}(t')$  is the generalized Kawasaki-Gunton projection operator depended on a structure of the relevant statistical operator (distribution function),  $\rho_{\text{rel}}(x^N; t')$ . By using Zubarev's NSO method [11], [12] and approach,  $\rho_{\text{rel}}(x^N; t')$  will be found from the extremum of the Renyi entropy at fixed values of observed values  $\langle \hat{P}_n(x) \rangle_\alpha^t$ , taking into account the normalization condition  $\langle I \rangle_{\alpha, \text{rel}}^t = 1$ , where the nonequilibrium average values are found respectively [5],

$$\langle \hat{P}_n(x) \rangle_\alpha^t = \hat{I}^\alpha(1, \dots, N) \hat{T}(1, \dots, N) \hat{P}_n \rho(x^N; t). \quad (6)$$

$\hat{I}^\alpha(1, \dots, N)$  has the following form for a system of  $N$  particles  $\hat{I}^\alpha(1, \dots, N) = \hat{I}^\alpha(1, \dots, \hat{I}^\alpha(N))$ ,  $\hat{I}^\alpha(j) = \hat{I}^\alpha(\vec{r}_j) \hat{I}^\alpha(\vec{p}_j)$  and defines operation of integration

$$\hat{I}^\alpha(x) f(x) = \int_{-\infty}^{\infty} f(x) d\mu_\alpha(x), \quad d\mu_\alpha(x) = \frac{|x|^\alpha}{\Gamma(\alpha)} dx. \quad (7)$$

The operator  $\hat{T}(1, \dots, N) = \hat{T}(1, \dots, \hat{T}(N))$  defines the operation  $\hat{T}(x_j) f(x_j) = \frac{1}{2} (f(\dots, x'_j - x_j, \dots) + f(\dots, x'_j + x_j, \dots))$ . Accordingly, the average value, which is calculated with the relevant distribution function, is defined as

$$\langle (\dots) \rangle_{\alpha, \text{rel}}^t = \hat{I}^\alpha(1, \dots, N) \hat{T}(1, \dots, N) (\dots) \rho_{\text{rel}}(x^N; t).$$

According to Ref. [12], from the extremum of the Renyi entropy functional

$$\begin{aligned} L_R(\rho') &= \frac{1}{1-q} \ln \hat{I}^\alpha(1, \dots, N) \hat{T}(1, \dots, N) (\rho'(t))^q \\ &\quad - \gamma \hat{I}^\alpha(1, \dots, N) \hat{T}(1, \dots, N) \rho'(t) \\ &\quad - \sum_n \int d\mu_\alpha(x) F_n(x; t) \hat{I}^\alpha(1, \dots, N) \hat{T}(1, \dots, N) \hat{P}_n(x) \rho'(t) \end{aligned}$$

at fixed values of observed values  $\langle \hat{P}_n(x) \rangle_\alpha^t$  and the condition of normalization  $\hat{I}^\alpha(1, \dots, N) \hat{T}(1, \dots, N) \rho'(t) = 1$ , the relevant distribution function takes the form

$$\rho_{\text{rel}}(t) = \frac{1}{Z_R(t)} \left[ 1 - \frac{q-1}{q} \beta \left( H - \sum_n \int d\mu_\alpha(x) F_n(x; t) \delta \hat{P}_n(x; t) \right) \right]^{\frac{1}{q-1}}, \quad (8)$$

where  $Z_R(t)$  is the partition function of the Renyi distribution, which is determined from the normalization condition and has the form

$$\begin{aligned} Z_R(t) &= \hat{I}^\alpha(1, \dots, N) \hat{T}(1, \dots, N) \\ &\quad \times \left[ 1 - \frac{q-1}{q} \beta \left( H - \sum_n \int d\mu_\alpha(x) F_n(x; t) \delta \hat{P}_n(x; t) \right) \right]^{\frac{1}{q-1}}. \end{aligned} \quad (9)$$

The Lagrangian multiplier  $\gamma$  is determined by the normalization condition. The parameters  $F_n(x; t)$  are determined from the self-consistency conditions

$$\langle \hat{P}_n(x) \rangle_\alpha^t = \langle \hat{P}_n(x) \rangle_{\alpha, \text{rel}}^t. \quad (10)$$

It is important to note that the relevant distribution function corresponded to the Gibbs entropy follows from (8) at  $q = 1$  [5]. In the general case of the parameters  $\langle \hat{P}_n(x) \rangle_\alpha^t$  of the reduced description of nonequilibrium processes according to (5) and (8), we get NSO in the form

$$\rho(t) = \rho_{\text{rel}}(t)$$

$$+ \sum_n \int d\mu_\alpha(x) \int_{-\infty}^t e^{\varepsilon(t'-t)} T(t, t') I_n(x; t') \rho_{\text{rel}}(t') \beta F_n^*(x; t') dt', \quad (11)$$

$$\text{where } F_n^*(x; t') = \frac{F_n(x; t')}{1 + \frac{q-1}{q} \sum_n \int d\mu_\alpha(x) F_n(x; t') \langle P_n(x) \rangle_\alpha^t},$$

$$I_n(x; t') = (1 - P(t)) \frac{1}{q} \psi^{-1}(t) iL_\alpha \hat{P}_n(x) \quad (12)$$

are the generalized flows,  $P(t)$  is the Mori projection operator [5], and the function  $\psi(t)$  has the following structure  $\psi(t) = 1 - \frac{q-1}{q} \sum_n \int d\mu_\alpha(x) F_n(x; t) P_n(x)$ .

By using the nonequilibrium statistical operator (11), we get the generalized transport equation for the parameters  $\langle \hat{P}_n(x) \rangle_\alpha^t$  of the reduced description,

$$\begin{aligned} \frac{\partial}{\partial t} \langle \hat{P}_n(x) \rangle_\alpha^t &= \langle iL_\alpha \hat{P}_n(x) \rangle_{\alpha, \text{rel}}^t \\ &\quad + \sum_{n'} \int d\mu_\alpha(x') \int_{-\infty}^t e^{\varepsilon(t'-t)} \varphi_{P_n P_{n'}}(x, x'; t, t') \beta F_{n'}^*(x'; t') dt', \end{aligned} \quad (13)$$

where

$$\begin{aligned} \varphi_{P_n P_{n'}}(x, x'; t, t') &= \hat{I}^\alpha(1, \dots, N) \hat{T}(1, \dots, N) \\ &\quad \times (iL_\alpha \hat{P}_n(x) T(t, t') I_{n'}(x'; t') \rho_{\text{rel}}(x^N; t')) \end{aligned} \quad (14)$$

are the generalized transport kernels (the memory functions), which describe dissipative processes in the system. To demonstrate the structure of the transport equations (13) and the transport kernels (14), we will consider, for example, diffusion processes. In the next section, we obtain generalized transport equations with fractional derivatives and consider a concrete example of diffusion processes of the particle in non-homogeneous media.

### III. GENERALIZED DIFFUSION EQUATIONS WITH FRACTIONAL DERIVATIVES

One of main parameters of the reduced description to describe the diffusion processes of the particles in non-homogeneous media with fractal structure is the nonequilibrium density of the particle numbers,  $\langle \hat{P}_n(x) \rangle_\alpha^t$ :  $n(\vec{r}; t) = \langle \hat{n}(\vec{r}) \rangle_\alpha^t$ , where  $\hat{n}(\vec{r}) = \sum_{j=1}^N \delta(\vec{r} - \vec{r}_j)$  is the microscopic density of the particles. The corresponding generalized diffusion equation for  $n(\vec{r}; t)$  can be obtained on base of Eqs. (8), (11), (13),

$$\frac{\partial \langle \hat{n}(\vec{r}) \rangle_\alpha^t}{\partial t} = \frac{\partial^\alpha}{\partial \vec{r}^\alpha} \cdot \int d\mu_\alpha(\vec{r}') \int_{-\infty}^t e^{\varepsilon(t'-t)} D_q(\vec{r}, \vec{r}'; t, t') \cdot \frac{\partial^\alpha \beta v^*(\vec{r}'; t')}{\partial \vec{r}'^\alpha} dt', \quad (15)$$

where

$$D_q(\vec{r}, \vec{r}'; t, t') = \langle \hat{v}(\vec{r}) T(t, t') \hat{v}(\vec{r}') \rangle_{\alpha, \text{rel}}^t \quad (16)$$

is the generalized coefficient diffusion of the particles within the Renyi statistics. Averaging in Eq. (16) is performed with the power-law Renyi distribution,

$$\rho_{rel}(t) = \frac{1}{Z_R(t)} \left( 1 - \frac{q-1}{q} \beta \left( H - \int d\mu_\alpha(\vec{r}) v^*(\vec{r}; t) \hat{n}(\vec{r}) \right)^{\frac{1}{q-1}} \right), \quad (17)$$

where

$$Z_R(t) = \hat{I}^\alpha(1, \dots, N) \hat{T}(1, \dots, N) \\ \times \left( 1 - \frac{q-1}{q} \beta \left( H - \int d\mu_\alpha(\vec{r}) v^*(\vec{r}; t) \hat{n}(\vec{r}) \right)^{\frac{1}{q-1}} \right) \quad (18)$$

is the partition function of the relevant distribution function,  $H$  is a Hamiltonian of the system,  $q$  is the Renyi parameter ( $0 < q < 1$ ).

Parameter  $v(\vec{r}; t)$  is the chemical potential of the particles, which is determined from the self-consistency condition,

$$\langle \hat{n}(\vec{r}) \rangle_\alpha^t = \langle \hat{n}(\vec{r}) \rangle_{\alpha, rel}^t. \quad (19)$$

$\beta = 1/k_B T$  ( $k_B$  is the Boltzmann constant),  $T$  is the equilibrium value of temperature,  $\hat{v}(\vec{r}) = \sum_{j=1}^N \vec{v}_j \delta(\vec{r} - \vec{r}_j)$  is the microscopic flux density of the particles. At  $q=1$ , the generalized diffusion equation within the Renyi statistics goes into the generalized diffusion equation within the Gibbs statistics with fractional derivatives. If  $q=1$  and  $\alpha=1$ , we obtain the generalized diffusion equation within the Gibbs statistics. In the Markov approximation, the generalized coefficient of diffusion in time and space has the form  $D_q(\vec{r}, \vec{r}'; t, t') \approx D_q \delta(t-t') \delta(\vec{r} - \vec{r}')$ . And by excluding the parameter  $v^*(\vec{r}'; t')$  via the self-consistency condition, we obtain the diffusion equation with fractional derivatives from Eq. (15)

$$\frac{\partial}{\partial t} \langle \hat{n}(\vec{r}) \rangle_\alpha^t = \sum_b D_q \frac{\partial^{2\alpha}}{\partial r^{2\alpha}} v^*(\vec{r}'; t'). \quad (20)$$

The generalized diffusion equation takes into account spatial fractality of the system and memory effects in the generalized coefficient of diffusion  $D_q(\vec{r}, \vec{r}'; t, t')$  within the Renyi statistics. Obviously, spatial fractality of system influences on transport processes of the particles that can show up as multifractal time with characteristic relaxation times. It is known that the nonequilibrium correlation functions  $D_q(\vec{r}, \vec{r}'; t, t')$  can not be exactly calculated, therefore the some approximations based on physical reasons are used. In the time interval  $-\infty \div t$ , ion transport processes in spatially non-homogeneous system can be characterized by a set of relaxation times that are associated with the nature of interaction between the particles and particles of media with fractal structure. To show the multifractal time in the generalized diffusion equation, we use the following approach for the generalized coefficient of particle diffusion

$$D_q(\vec{r}, \vec{r}'; t, t') = W(t, t') \bar{D}_q(\vec{r}, \vec{r}'), \quad (21)$$

where  $W(t, t')$  can be defined as the time memory function. In view of this, Eq. (15) can be represented as

$$\frac{\partial}{\partial t} \langle \hat{n}(\vec{r}) \rangle_\alpha^t = \int_{-\infty}^t e^{\epsilon(t'-t)} W(t, t') \Psi(\vec{r}; t') dt', \quad (22)$$

where

$$\Psi(\vec{r}; t') = \int d\mu_\alpha(\vec{r}') \frac{\partial^\alpha}{\partial \vec{r}'^\alpha} \cdot \bar{D}_q(\vec{r}, \vec{r}') \cdot \frac{\partial^\alpha}{\partial \vec{r}'^\alpha} \beta v^*(\vec{r}'; t'). \quad (23)$$

Further we apply the Fourier transform to Eq. (22), and as a result we get in frequency representation

$$i\omega n(\vec{r}; \omega) = W(\omega) \Psi(\vec{r}; \omega). \quad (24)$$

We can represent the frequency dependence of the memory function in the following form

$$W(\omega) = \frac{(i\omega)^{1-\xi}}{1+i\omega\tau}, \quad 0 < \xi \leq 1, \quad (25)$$

where the introduced relaxation time  $\tau_a$  characterizes the particles transport processes in the system. Then Eq. (24) can be represented as

$$(1+i\omega\tau)i\omega n(\vec{r}; \omega) = (i\omega)^{1-\xi} \Psi(\vec{r}; \omega). \quad (26)$$

Further we use the Fourier transform to fractional derivatives of functions,

$$(_0 D_t^{1-\xi} f(t); i\omega) = (i\omega)^{1-\xi} L(f(t); i\omega). \quad (27)$$

By using it, the inverse transformation of Eq. (26) to time representation gives the Cattaneo-type generalized diffusion equation with taking into account spatial fractality,

$$\tau \frac{\partial^2}{\partial t^2} n(\vec{r}; t) + \frac{\partial}{\partial t} n(\vec{r}; t) = {}_0 D_t^{1-\xi} \Psi(\vec{r}; t) = \frac{\partial^{1-\xi}}{\partial t^{1-\xi}} \Psi(\vec{r}; t), \quad (28)$$

which is the new Cattaneo-type generalized equation within the Renyi statistics with multifractal time and spatial fractality. At  $q=1$  from Eq. (29), we get the Cattaneo-type generalized equation within the Gibbs statistics with multifractal time and spatial fractality,

$$\tau \frac{\partial^2}{\partial t^2} n(\vec{r}; t) + \frac{\partial}{\partial t} n(\vec{r}; t) = {}_0 D_t^{1-\xi} \int d\mu_\alpha(\vec{r}') \frac{\partial^\alpha}{\partial \vec{r}'^\alpha} \cdot \bar{D}(\vec{r}, \vec{r}') \\ \cdot \frac{\partial^\alpha}{\partial \vec{r}'^\alpha} \beta v(\vec{r}'; t), \quad (29)$$

Eqs. (28), (29) contain significant spatial non-homogeneity in  $\bar{D}_q(\vec{r}, \vec{r}')$ . If we neglect spatial non-homogeneity,

$$\bar{D}_q(\vec{r}, \vec{r}') = \bar{D}_q \delta(\vec{r} - \vec{r}'), \quad (30)$$

we get the Cattaneo-type diffusion equation with fractality of space-time and the constant coefficients of the diffusion within the Renyi statistics,

$$\tau \frac{\partial^2}{\partial t^2} n(\vec{r}; t) + \frac{\partial}{\partial t} n(\vec{r}; t) = {}_0 D_t^{1-\xi} \bar{D}_q \frac{\partial^{2\alpha}}{\partial \vec{r}^{2\alpha}} \beta v^*(\vec{r}; t), \quad (31)$$

At  $q=1$ , we get the Cattaneo-type diffusion equation with fractality of space-time and the constant coefficients of the diffusion within the Gibbs statistics,

$$\tau \frac{\partial^2}{\partial t^2} n(\vec{r}; t) + \frac{\partial}{\partial t} n(\vec{r}; t) = {}_0 D_t^{1-\xi} \sum_b \bar{D} \frac{\partial^{2\alpha}}{\partial \vec{r}^{2\alpha}} v(\vec{r}; t), \quad (32)$$

It should be noted that if we put  $\alpha=1$  in Eqs. (31), (32), i.e. we neglect spatial fractality, we get the Cattaneo-type diffusion equations, which were obtained in Ref. [18],

$$\tau \frac{\partial^2}{\partial t^2} n(\vec{r}; t) + \frac{\partial}{\partial t} n(\vec{r}; t) = {}_0 D_t^{1-\xi} \bar{D} \frac{\partial^2}{\partial \vec{r}^2} v(\vec{r}; t). \quad (33)$$

At  $\tau = 0$ , we get an important particular case — the generalized diffusion equation of particles with taking into account fractality of space-time,

$$\frac{\partial}{\partial t} n(\vec{r}; t) = {}_0 D_t^{1-\xi} \int d\mu_\alpha(\vec{r}') \frac{\partial^\alpha}{\partial \vec{r}'^\alpha} \cdot \bar{D}_q(\vec{r}, \vec{r}') \cdot \frac{\partial^\alpha}{\partial \vec{r}'^\alpha} \beta v^*(\vec{r}'; t), \quad (34)$$

and by neglecting spatial non-homogeneity of the diffusion coefficients  $\bar{D}_q(\vec{r}, \vec{r}')$ , we also get the diffusion equation with the constant coefficients of the diffusion with the fractional derivatives within the Renyi statistics,

$$\frac{\partial}{\partial t} n(\vec{r}; t) = {}_0 D_t^{1-\xi} \bar{D}_q \frac{\partial^{2\alpha}}{\partial \vec{r}^{2\alpha}} \beta v^*(\vec{r}; t), \quad (35)$$

At  $\alpha = 1$ ,  $\tau = 0$ , we get the diffusion equation with the constant coefficients of the diffusion without spatial fractality within the Renyi statistics

$$\frac{\partial}{\partial t} n(\vec{r}; t) = {}_0 D_t^{1-\xi} \bar{D}_q \frac{\partial^2}{\partial \vec{r}^2} \beta v^*(\vec{r}; t), \quad (36)$$

At  $\alpha = 1$ ,  $\tau = 0$ ,  $q = 1$ ,  $\xi = 1$ , we get the usual diffusion equation for the particles within the Gibbs statistics,

$$\frac{\partial}{\partial t} n(\vec{r}; t) = \bar{D} \frac{\partial^2}{\partial \vec{r}^2} \beta v(\vec{r}; t). \quad (37)$$

Let us consider another model of the memory function

$$W(\omega) = \frac{(i\omega)^{1-\xi}}{1 + (i\omega\tau)^{\gamma-1}}, \quad (38)$$

then in frequency representation we get

$$(1 + (i\omega\tau)^{\gamma-1}) \hat{n}(\vec{r}; \omega) = (i\omega)^{1-\xi} \Psi(\vec{r}; \omega). \quad (39)$$

By using Eq. (27) and inverse transformation of Eq. (39) to the time  $t$ , we get the generalized Cattaneo-type diffusion equation with taking into account multifractal time and spatial fractality.

#### IV. CONCLUSION

We presented the general approach for obtaining the generalized transport equations with the fractional derivatives by using the Liouville equation with the fractional derivatives [14] for a system of classical particles and Zubarev's NSO method within the Renyi statistics [5]. In this approach, the new non-Markov equations of diffusion of the particles in a spatially non-homogeneous medium with a fractal structure are obtained.

By using approaches for the memory functions and fractional calculus [1]–[5], the generalized Cattaneo-type diffusion equations with taking into account fractality of space-time are obtained. It is considered the different models for the frequency dependent memory functions, which lead to the known diffusion equations with the fractality of space-time and their generalizations.

#### REFERENCES

- [1] K. B. Oldham and J. Spanier, *The Fractional Calculus: Theory and Applications of Differentiation and Integration to Arbitrary Order*. Dover Books on Mathematics; Dover Publications, 2006.
- [2] S. G. Samko, A. A. Kilbas, and O. I. Marichev, *Fractional Integrals and Derivatives: Theory and Applications*. Gordon and Breach Science Publishers, 1st ed., 1993.
- [3] I. Podlubny and V. T. E. Kenneth, *Fractional Differential Equations: An Introduction to Fractional Derivatives, Fractional Differential Equations, to Methods of Their Solution and Some of Their Applications*. Academic Press, 1st ed., 1998.
- [4] V. V. Uchaikin, *Fractional Derivatives Method*. Artishock-Press, Uljanovsk, 2008.
- [5] P. Kostrobij, B. Markovich, O. Viznovych, and M. Tokarchuk, "Generalized diffusion equation with fractional derivatives within Renyi statistics." *Journal of Mathematical Physics*, vol. 57, pp. 093301, 2016.
- [6] P. Kostrobij, B. Markovich, O. Viznovych, and M. Tokarchuk, "Generalized electrodiffusion equation with fractality of space-time." *Mathematical Modeling and Computing*, vol. 3, n. 2, pp. 163–172, 2016.
- [7] R. R. Nigmatullin, "To the Theoretical Explanation of the ‘Universal Response’." *Physica Status Solidi (b)*, vol. 123, pp. 739–745, 1984.
- [8] R. R. Nigmatullin, "On the Theory of Relaxation for Systems with ‘Remnant’ Memory." *Physica Status Solidi (b)*, vol. 124, pp. 389–393 1984.
- [9] R. R. Nigmatullin, "The realization of the generalized transfer equation in a medium with fractal geometry." *Physica Status Solidi (b)*, vol. 133, pp. 425–430, 1986.
- [10] R. R. Nigmatullin, "Fractional integral and its physical interpretation." *Theoretical and Mathematical Physics*, vol. 90, pp. 242–251, 1992.
- [11] D. N. Zubarev, V. G. Morozov, and G. Röpke, *Statistical mechanics of nonequilibrium processes*. Fizmatlit, Vol. 1, 2, 2002.
- [12] B. B. Markiv, R. M. Tokarchuk, P. P. Kostrobij, and M. V. Tokarchuk, "Nonequilibrium statistical operator method in Renyi statistics." *Physica A: Statistical Mechanics and its Applications*, vol. 390, pp. 785–791, 2011.
- [13] V. E. Tarasov, *Fractional Dynamics: Applications of Fractional Calculus to Dynamics of Particles, Field and Media*. Springer, New York, 2011.
- [14] V. E. Tarasov, "Transport equations from Liouville equations for fractional systems." *International Journal of Modern Physics B*, vol. 20, pp. 341–353, 2006.
- [15] K. Cottrill-Shepherd and M. Naber, "Fractional differential forms." *Journal of Mathematical Physics*, vol. 42, pp. 2203–2212, 2001.
- [16] F. Mainardi, *In Fractals and Fractional Calculus in Continuum Mechanics*; A. Carpinteri, F. Mainardi, Eds.; Springer Vienna: Vienna, pp. 291–348, 1997.
- [17] M. Caputo and F. Mainardi, "A new dissipation model based on memory mechanism." *Pure and Applied Geophysics*, vol. 91, pp. 134–147, 1971.
- [18] A. Compte and R. Metzler, "The generalized Cattaneo equation for the description of anomalous transport processes." *Journal of Physics A: Mathematical and General*, vol. 30, pp. 7277–7289, 1997.

# Defining of Lyapunov Functions for the Generalized Linear Dynamical Object

Roman Voliansky<sup>1</sup>, Oleksander Sadovoi<sup>1</sup>, Yuliia Sakhina<sup>1</sup>, Nina Volianska<sup>2</sup>

1. Department of Electrotechnic and Electromechanic, Dniprovske State Technical University, Ukraine, Kamyanske,  
2 Dniprobusdovska str., email: [voliansky@ua.fm](mailto:voliansky@ua.fm)

2. Heat Transfer Department, Dniprovske State Technical University, Ukraine, Kamyanske,  
2 Dniprobusdovska str., email: [ninanin@ua.fm](mailto:ninanin@ua.fm)

**Abstract:** The paper deals with the developing method of determination of Lyapunov functions for a generalized linear dynamical object. This method is based on the shift and the rotation coordinate transformation which translates a motion of the considered object in a new virtual state space. One can perform such sort of transformation by using partial fraction decomposition. It is easy to define Lyapunov function in a new state space and then do inverse transformation. The proposed method can be used for a determination of signed Lyapunov functions, which can be used as basis for the analysis of system dynamic and the synthesis of desired motions.

**Keywords:** stability analysis, Lyapunov function, coordinate transformation, dynamical system, partial fraction decomposition.

## I. INTRODUCTION

Nowadays Lyapunov functions are widely used to solve different control problems. One can find these functions very usable while synthesis and analysis problems are being solved. The great interest to Lyapunov functions can be explained by their unique properties and strong mathematical background which allows getting mathematically valid results. One can easily use these results while optimization problems are formulated for various dynamical systems.

Although one can find a lot of publications in recent scientific periodicals which describe definitions of non-quadratic Lyapunov functions [1], quadratic forms are still commonly used for defining candidate to Lyapunov function [2-4]. This fact can be simply explained by physical meaning of these quadratic functions which have an energetic background and show redundant energy of dynamical object.

Due to Lyapunov's theorem about stability of motion corresponding Lyapunov function must satisfy Sylvester criterion [5]. Since, there is an infinity number of quadratic functions which satisfy this criterion, their definition is a nontrivial problem of the control theory. This problem is solved by using Riccati and/or Lyapunov equations, which in common case depend on some cost function and can be solved only numerically [6,7].

In order to avoid above-mentioned drawbacks of Lyapunov function's we suggest define them by developing analytical method for defining form and coefficients of Lyapunov

function only as functions on parameters of the considered object.

Our paper is organized as follows: first of all we consider the transformation of a generalized linear object into parallel form. Secondly we define Lyapunov function for the transformed object. Thirdly, we perform inverse transformation and write down Lyapunov function which depends only on parameters and coordinates of dynamical object. Lastly, we show the example of using proposed approach and make a conclusion.

## II. USAGE OF PARALLEL MODEL FOR SIMULATION AND ANALYSIS OF DYNAMICAL SYSTEM

### A. Representation of object's dynamic in parallel way

Let us consider a linear single-input dynamical object which dynamic is given as follows

$$sx_j = \sum_{i=1}^n b_{ij}x_i + m_n U, \quad (1)$$

where  $s = d/dt$  is a derivative operator,  $x_i, x_j$  are state variables,  $U$  is a control input,  $b_{ij}, m_n$  are coefficients,  $n$  is an order of dynamical object.

Equations (1) can be rewriting into matrix form

$$s\mathbf{X} = \mathbf{B}\mathbf{X} + \mathbf{M}U, \quad (2)$$

where  $\mathbf{X} = (x_1 \ x_2 \ \dots \ x_n)^T$  is a state space vector,  $\mathbf{M} = (0 \ 0 \ \dots \ m_n)^T$  is a  $n$ -th sized vector of input coefficients, and  $\mathbf{B}$  is a  $n$ -th sized square matrix of coefficients

$$\mathbf{B} = \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \dots & b_{nn} \end{pmatrix}. \quad (3)$$

One can find a matrix transfer function of the control object (1) in an easy way by using equation (2)

$$\mathbf{W}(s) = (s\mathbf{E} - \mathbf{B})^{-1} \mathbf{M}, \quad (4)$$

here  $\mathbf{E}$  is an identity matrix.

We assume that this matrix  $\mathbf{B}$  has only real eigenvalues. In this case its characteristic polynomial can be written thus

$$D(s) = \det(s\mathbf{E} - \mathbf{B}) = \prod_{i=1}^n (s + \lambda_i), \quad (5)$$

where  $\lambda_i$  are the eigenvalues of matrix  $\mathbf{B}$ .

Now we suggest to simplify transfer function (4) in the following way

$$\mathbf{W}(s) = \prod_{i=1}^n \frac{1}{(s + \lambda_i)} \mathbf{A}, \quad (6)$$

where

$$\mathbf{A} = \text{adj}(s\mathbf{E} - \mathbf{B})\mathbf{M}, \quad (7)$$

here  $\text{adj}(s\mathbf{E} - \mathbf{B})$  is the matrix which is adjunct to matrix  $s\mathbf{E} - \mathbf{B}$ .

It is clear that first cofactor of expression (6) contain n-fold multiplication of elementary fractions. This multiplication can be replaced with a sum of some elementary fractions as follows [8]

$$\prod_{i=1}^n \frac{1}{(s + \lambda_i)} = \sum_{i=1}^n \frac{\alpha_i}{(s + \lambda_i)}, \quad (8)$$

where

$$\begin{aligned} \sum_{i=1}^n \alpha_i &= 0; \quad \sum_{i=1}^n \alpha_i \left( \sum_{j=1}^{i-1} \lambda_j + \sum_{j=i+1}^n \lambda_j \right) = 0 \\ &\vdots \\ \sum_{i=1}^n \alpha_i \left( \prod_{j=1}^{i-1} \lambda_j \cdots \prod_{j=i+1}^n \lambda_j \right) &= 1. \end{aligned} \quad (9)$$

Expression (9) is obtained by using only characteristic polynomial (5) and it is independent of selected component of matrix  $\mathbf{A}$ . This fact allows to rewrite matrix transfer function (6) thus

$$\mathbf{W}(s) = \sum_{i=1}^n \frac{\alpha_i}{(s + \lambda_i)} \mathbf{A}. \quad (10)$$

Thereby, we replace the series transfer function (6) where the calculation can be performed only by using consecutive calculations with the parallel one (10) which can be calculated in a parallel way. One of the benefits of such an approach is increasing of the calculation speed while parallel simulation is implemented.

### B. Direct and inverse coordinate transformations

Apart from above-mentioned calculation advantage, the proposed approach has significant methodological values. One can find these methodological benefits while performing stability analysis and considering energy transformation. In this case the proposed approach allows us to perform some coordinate transformation from normal phase space to some virtual one and in such a way simplify Lyapunov function.

Let us consider these transformations in detail.

First of all, we define square matrix  $\mathbf{Y}$  as follows

$$\mathbf{Y} = \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{nn} \end{pmatrix} \quad (11)$$

and set the following interrelation

$$x_j = \sum_{i=1}^n y_{ji}, \quad (12)$$

where

$$y_{ji} = W_{ji}(s)U, \quad (13)$$

here  $a_i$  is the i-th component of matrix  $\mathbf{A}$ ,

$$W_{ji}(s) = \alpha_j a_i / (s + \lambda_j). \quad (14)$$

Expression (12) allows us make the following statement.

**Statement 1.** For state variable  $x_j$  inverse coordinate transformation from virtual phase space into normal is performed by summing all relevant virtual space variable  $y_{ji}$ .

Now we consider determination of  $y_{ji}$  coordinates while direct transformation is being performed.

Let us take into account equation (15) and complete equation (12) with its first n-1 derivatives. In such a way we get a n-th order system of linear equations with n unknown variables  $y_{ji}$

$$L_f^k x_j = \sum_{i=1}^n L_f^k y_{ji}, \quad k = 0, \dots, n-1, \quad (15)$$

where  $L_f^k x_j, L_f^k y_{ji}$  are k-th order Lie derivatives.

Solution of this system for unknown virtual state variable  $y_{ji}$  allows us define them in such a way

$$y_{ji} = \sum_{k=1}^n \gamma_{kji} x_k + \kappa_{ji} U, \quad (16)$$

where  $\gamma_{kji}, \kappa_{ji}$  are some numbers.

Expressions (12) and (16) allow us claim the following.

**Statement 2.** Direct and inverse transformations are described with simple algebraic expressions and it is defined with some family of shift and rotation transformations.

That is why the above-given transformation can be used for simplification of a dynamical system description and performing some actions like stability analysis.

### C. Stability analysis

It is clearly understood that one can use expressions (12) and (16) for stability analysis. This analysis after performing transformation (8) comes down to analysis of stability every transfer function (14). This fact allows us formulate the following statement.

**Statement 3.** Considered dynamical object has stable dynamics on condition that every parallel channels has stable dynamics as well. In this case we can claim that all components of vector  $\mathbf{Y}$  are bounded.

One can perform stability analysis for each channel in a different way. The simplest one is analysis of  $\lambda_i$  eigenvalues. The more complex one is based on usage of Lyapunov functions. In spite of its complexity Lyapunov function allows us not only to do stability analysis but consider energy conversion while dynamical object is operating, also define algorithms and structure of controller for the considered object.

#### D. Lyapunov function dedetrmination

The simplest Lyapunov function is the following quadratic expression

$$V_{ji} = k_{ji} y_{ji}^2, \quad (17)$$

where  $k_{ji}$  is a positive number.

So, while analysis of stability is being performed, one can use expression (12) and determine the following Lyapunov function

$$V_j = \sum_{i=1}^n k_{ji} y_{ji}^2. \quad (18)$$

The function (18) can be written down as matrix expression

$$V_j = \mathbf{Y}_j \mathbf{K}_j \mathbf{Y}_j^T, \quad (19)$$

where

$$\mathbf{Y}_j = (y_1 \ y_2 \ \dots \ y_n), \quad (20)$$

$$\mathbf{K}_j = \begin{pmatrix} k_{j1} & 0 & \dots & 0 \\ 0 & k_{j2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & k_{jn} \end{pmatrix}. \quad (21)$$

Quadratic form (19) depends only on diagonal matrix (21). It is quite clear that determinant of matrix (21) and its diagonal minors can be easily defined thus

$$\det(\mathbf{K}_j) = \prod_{i=1}^n k_{ji} \quad (22)$$

$$M_k = \prod_{i=1}^k k_{ji}, \quad k = 1, \dots, n. \quad (23)$$

Analysis of expressions (22) and (23) allows us to formulate the following statement.

**Statement 4.** According to Sylvester criterion [6] Lyapunov function (19) is positive if  $k_{ji}$  coefficients are positive.

Positiveness of j-th Lyapunov function (19) means stability of j-th channel dynamic.

One can substitutes interrelation (16) info function (18) and write down following expression for Lyapunov function in real state variables

$$V_j = \sum_{i=1}^n k_{ji} \left( \sum_{k=1}^n \gamma_{kji} x_k + \kappa_{ji} U \right)^2. \quad (24)$$

Lyapunov function (24) is a positive function as well due to positivenes of function (18).

One can open brackets in function (24) and transform it into modification of well-known quadratic Lyapunov function

$$V_j = \mathbf{Z} \mathbf{W} \mathbf{Z}^T, \quad (25)$$

where

$$\mathbf{Z} = (x_1 \ x_2 \ \dots \ x_n \ U), \quad (26)$$

$$\mathbf{W} = \begin{pmatrix} w_{11} & \dots & w_{1(n+1)} \\ \vdots & \ddots & \vdots \\ w_{(n+1)1} & \dots & w_{(n+1)(n+1)} \end{pmatrix}, \quad (27)$$

here

$$w_{ii} = k_{ij} \gamma_{kji}, \quad i = 1, \dots, n; \quad w_{i(n+1)} = k_{ij} \kappa_{ji};$$

$$w_{ij} = 2k_{ji} \gamma_{kji} \gamma_{mji}, \quad i, m = 1, \dots, n; \quad (28)$$

$$w_{ij} = 2k_{ji} \gamma_{kji} \kappa_{ji}, \quad (i = n) \text{ or } (j = n)$$

The function (25) is a j-th component of matrix Lyapunov function

$$\mathbf{V} = (V_1 \ V_2 \ \dots \ V_n), \quad (29)$$

which describes redundant energy stored in each channel.

One can use Lyapunov function defined in such a way for stability analysis and design of closed-loop control system. Now let us consider example of defining Lyapunov functions for the speed and current loops of DC motor.

### III. EXAMPLE USE FOR PROPOSED APPROACH

Let us consider a dynamic of DC motor given by the following equations

$$sx_1 = b_{12}x_2; sx_2 = b_{21}x_1 + b_{22}x_2 + m_2U, \quad (30)$$

where coefficients  $a_{ij}$  are defined thus

$$b_{12} = \frac{I}{T_m}; b_{21} = b_{22} = -\frac{I}{T_e}; m_2 = \frac{1}{T_e}; T_m = \frac{JR}{c^2}; T_e = \frac{L}{R}, \quad (31)$$

here J is a rotor inertia, R is a armature resistanc, c is a back-emf constant, L is an armature inductance,  $y_1, y_2$  are DC rotor speed and current respectively.

We assume that the rotor inertia has significant value and the following condition is true

$$T_m > 4T_e. \quad (32)$$

In this case both of eigenvalues of the characteristic polynomial

$$D(s) = s^2 - b_{22}s - b_{12}b_{21}. \quad (33)$$

are negative

$$\lambda_{1,2} = 0.5b_{22} \pm 0.5\sqrt{4b_{12}b_{21} + b_{22}^2}. \quad (34)$$

This fact allows us represent dynamic DC motor in form (7)

$$\mathbf{W}(s) = \begin{pmatrix} x_1(s)/U(s) \\ x_2(s)/U(s) \end{pmatrix} = \frac{\mathbf{A}}{s^2 - b_{22}s - b_{12}b_{21}}, \quad (35)$$

where

$$\mathbf{A} = (m_2 b_{12} \ m_2 s)^T \quad (36)$$

or

$$\mathbf{W}(s) = \begin{pmatrix} \frac{m_2 b_{12}}{s^2 - b_{22}s - b_{12}b_{21}} & \frac{m_2 s}{s^2 - b_{22}s - b_{12}b_{21}} \end{pmatrix}^T. \quad (37)$$

Now we transform transfer function (35) into the form (10)

$$\frac{\mathbf{A}}{s^2 - b_{22}s - b_{12}b_{21}} = \frac{\mathbf{a1}}{s - \lambda_1} + \frac{\mathbf{a2}}{s - \lambda_2}, \quad (38)$$

where

$$\mathbf{a1} = \begin{pmatrix} b_{12}m_2 & m_2\lambda_1 \\ \lambda_1 - \lambda_2 & \lambda_1 - \lambda_2 \end{pmatrix}^T; \mathbf{a2} = \begin{pmatrix} -b_{12}m_2 & -m_2\lambda_2 \\ \lambda_1 - \lambda_2 & \lambda_1 - \lambda_2 \end{pmatrix}^T. \quad (39)$$

The matrices (39) allow us write the following equations

$$\begin{aligned} sy_{11} &= \lambda_1 y_{11} + \alpha_1 I_1 U; & sy_{12} &= \lambda_2 y_{12} + \alpha_2 I_1 U; \\ sy_{21} &= \lambda_1 y_{21} + \alpha_1 I_2 U; & sy_{22} &= \lambda_2 y_{22} + \alpha_2 I_2 U. \end{aligned} \quad (40)$$

These equations allow us rewrite interrelations (12) and (16) between real and virtual coordinates if the output variable is the variable  $x_1$

$$x_1 = y_{11} + y_{12}; x_2 = \frac{\lambda_1}{b_{12}} y_{11} + \frac{\lambda_2}{b_{12}} y_{12} + \frac{\alpha_1 I + \alpha_2 U}{b_{12}} \quad (41)$$

and if the output variable is the variable  $x_2$

$$x_2 = y_{21} + y_{22}; \\ x_1 = \frac{\lambda_1 - b_{22}}{b_{21}} y_{21} + \frac{\lambda_2 - b_{22}}{b_{21}} y_{22} + \frac{\alpha_1 I + \alpha_2 U - m_2}{b_{21}} \quad (42)$$

We consider expression (41) and (42) as inverse coordinate transformation. The expressions for the direct transformation can be obtained as solution equations (41) and (42) for state variables  $y_{11}, y_{12}, y_{21}, y_{22}$

$$y_{11} = \frac{-\lambda_2 x_1 - x_2 b_{12} + U(\alpha_1 I + \alpha_2 U)}{\lambda_1 - \lambda_2}; \quad (43) \\ y_{12} = \frac{-x_2 b_{12} + \lambda_1 x_1 + U(\alpha_1 I + \alpha_2 U)}{\lambda_1 - \lambda_2} \\ y_{21} = \frac{b_{21} x_1 + (-\lambda_2 + b_{22}) x_2 - (\alpha_1 I + \alpha_2 U - m_2) U}{\lambda_1 - \lambda_2}; \quad (44) \\ y_{22} = \frac{-b_{21} x_1 + (\lambda_1 - b_{22}) x_2 + (\alpha_1 I + \alpha_2 U - m_2) U}{\lambda_1 - \lambda_2}.$$

One can use coordinates (43) and (44) to determine Lyapunov functions. The simplest ones can be written down for the speed loop

$$V_I = y_{11}^2 + y_{12}^2 = w_{11} x_1^2 + 2w_{12} x_1 x_2 + w_{22} x_2^2 + \\ + 2w_{13} x_1 U + 2w_{23} x_2 U + w_{33} U^2, \quad (45)$$

where

$$w_{11} = \frac{\lambda_1^2 + \lambda_2^2}{(\lambda_1 - \lambda_2)^2}; w_{12} = \frac{-b_{12}(\lambda_1 + \lambda_2)}{(\lambda_1 - \lambda_2)^2}; \\ w_{22} = \frac{2b_{12}^2}{(\lambda_1 - \lambda_2)^2}; w_{23} = \frac{2b_{12}(\alpha_1 I + \alpha_2 U)}{(\lambda_1 - \lambda_2)^2}; \quad (46) \\ w_{33} = \frac{2(\alpha_1 I + \alpha_2 U)^2}{(\lambda_1 - \lambda_2)^2}; w_{13} = \frac{2(\lambda_1 + \lambda_2)(\alpha_1 I + \alpha_2 U)}{(\lambda_1 - \lambda_2)^2}.$$

Lyapunov function for the current loop can be defined in a similar way but it has different coefficients

$$w_{11} = \frac{2b_{21}^2}{(\lambda_1 - \lambda_2)^2}; w_{22} = \frac{(\lambda_2 - b_{22})^2 + (\lambda_1 - b_{22})^2}{(\lambda_1 - \lambda_2)^2}; \quad (47) \\ w_{12} = -\frac{b_{21}(\lambda_2 - b_{22})}{(\lambda_1 - \lambda_2)^2} - \frac{b_{21}(\lambda_1 - b_{22})}{(\lambda_1 - \lambda_2)^2}; \\ w_{13} = \frac{-2b_{21}(\alpha_1 I + \alpha_2 U - m_2)}{(\lambda_1 - \lambda_2)^2}; w_{33} = \frac{2(\alpha_1 I + \alpha_2 U - m_2)^2}{(\lambda_1 - \lambda_2)^2}; \\ w_{23} = \frac{(\lambda_1 + \lambda_2 - 2b_{22})(\alpha_1 I + \alpha_2 U - m_2)}{(\lambda_1 - \lambda_2)^2}.$$

If one takes into account matrices (39) and performs analysis of coefficients (46) and (47), it still can be possible define that coefficients  $w_{13}, w_{23}, w_{33}$  in expression (46) are equal

to zero and coefficients  $w_{13}, w_{23}, w_{33}$  in expression (47) can be simplified as follows

$$w_{11} = \frac{2b_{21}^2}{(\lambda_1 - \lambda_2)^2}; w_{22} = \frac{(\lambda_2 - b_{22})^2 + (\lambda_1 - b_{22})^2}{(\lambda_1 - \lambda_2)^2}; \\ w_{12} = -\frac{b_{21}(\lambda_2 - b_{22})}{(\lambda_1 - \lambda_2)^2} - \frac{b_{21}(\lambda_1 - b_{22})}{(\lambda_1 - \lambda_2)^2}; w_{13} = \frac{-2m_2 b_{21}}{(\lambda_1 - \lambda_2)^2}; \quad (48) \\ w_{33} = \frac{2m_2^2}{(\lambda_1 - \lambda_2)^2}; w_{23} = \frac{-m_2(\lambda_1 + \lambda_2 - 2b_{22})}{(\lambda_1 - \lambda_2)^2}.$$

This fact allows us formulate following statement.

**Statement 5.** One should define Lyapunov function (25) in an extended  $n+1$ -th order space state with space vector (26) if output variable is described with the differential equation which contains control input. It can be defined in a normal  $n$ -th order state space otherwise.

#### IV. CONCLUSION

The proposed approach based on decomposition of transfer function of linear dynamical object with elementary fractions can be used for simulation of considered object, the study of its dynamics and synthesis of its control system. This approach simplifies mathematical model of a linear object and transform this model into some virtual state space. The mentioned transformation allows us define Lyapunov function in an easy way. This function can be defined for both object and linear closed-loop control system. In this case its coefficients depend only on parameters of dynamical system.

#### REFERENCES

- [1] Hu Tingshu, et.al Non-quadratic Lyapunov functions for performance analysis of saturated systems, *Proceedings of 44<sup>th</sup> IEEE conference on Decision and Control*, pp.8106-8111, 2005
- [2] T.L. Vu , K. Turitsyn Lyapunov Functions Family Approach to Transient Stability Assessment *IEEE Transactions on Power Systems*, Volume: 31, Issue: 2, pp 1269 — 1277, 2015
- [3] A. Bacciotti, L. Rosier, Liapunov functions and stability in control theory. Berlin: Springer, 2005, 237p.
- [4] M. Malisoff, F. Mazenc, Constructions of strict Lyapunov Functions. London: Springer, 2009, 386p.
- [5] G.Giorgio, Various proofs of the Sylvestr criterion for quadratic forms, *Journal of Mathematical Researches*, Vol. 9(6), pp.55-66, 2017
- [6] T. Penzi, Numerical solution of generalized Lyapunov equations, *Advanced in Comp. Math.*, Vol.8 pp.33-48, 1898
- [7] W.F. Arnold and A.J.Laub Generaliztion eigenproblem algorithms and software for algebraic Riccati equations, *Proceeding of the IEEE*, Vol. 72, pp1746-1754, 1984
- [8] R. Witula and D. Slota Partial fractions decompositions of some rational functions, *Appl. Math. Comput.*, Vol. 197, pp.328-336, 2008.

# Mathematical Modeling of Deformation-Relaxation Processes under Phase Transition

Yaroslav Sokolovskyy<sup>1</sup>, Iryna Boretska<sup>2</sup>, Svitlana Yatsyshyn<sup>3</sup>, Yaroslav Kaspryshyn<sup>4</sup>

IT Department, Ukrainian National Forestry University, UKRAINE, Lviv, 103 Chuprynska street, e-mail: [sokolowskyy@yahoo.com](mailto:sokolowskyy@yahoo.com)<sup>1</sup>, [iryna.boretska@gmail.com](mailto:iryna.boretska@gmail.com)<sup>2</sup>, [svitlana0981@gmail.com](mailto:svitlana0981@gmail.com)<sup>3</sup>, [kaspryshyn.ya@gmail.com](mailto:kaspryshyn.ya@gmail.com)<sup>4</sup>.

**Abstract:** This paper presents the mathematical modeling of deformation-relaxation processes under phase transition.

**Keywords:** phase transition, capillary-porous materials, viscoelastic stresses, deformation-relaxation processes, heat-mass-exchange processes.

## I. INTRODUCTION

During the process of drying capillary-porous materials, the zone of evaporation of moisture is deepening to the middle of the material. The presence of moving boundary of phase transformations at the interface between phases with different thermophysical and mechanical characteristics considerably complicates the mathematical models of deformation-relaxation and heat-mass-exchange processes during the drying of capillary-porous materials. The modeling of heat-mass-exchange process with phase transitions in the drying process is diminished to the solving of Stefan's problem, which is the most complicated even for minor changes of material's density in the evaporation zone. However, the evaporation of water causes a change of its volume of almost a thousand times, and the removal of the vapor-gas mixture from the evaporation zone requires significant energy expenditures. With the deepening of the evaporation zone in the volume of the drying material, appears a significant increase of a pressure near the evaporation front. Therefore, taking into account the energy consumption of the steam movement kinetics and the convective heat transfer to the evaporation zones is included into different approaches of representing the evaporation zone model. If the material being drying is characterized by rheological properties, the part of energy associated with irreversible deformations is dissipated in the material. Therefore, the equation of the balance of dysplasia energy on the surface of the phase transition makes it possible to formulate conditions for the zone of deepening of the evaporation, taking into account the storage of irreversible deformations at phase transitions.

The development of mathematical models of viscoelastic capillary-porous materials in the process of drying is based on models of consequential creep [1,2]. However, such models describe the rheological behavior of different environments for a continuous history of deformation. In case the process of deformation of capillary-porous environments has implicit phase transitions and on their boundaries it is characterized by ruptures, it is necessary to take into account the influence of the previous history of the load on the phase transition on the further development of the stress-deformed

state of the environment. For homogeneous viscoelastic environments that have changed the phase transition during deformation process, a hypothesis is used to preserve viscoelastic stresses at the transition boundary [3]. This hypothesis is based on the using the main theorems of the theory of viscoelasticity, in particular Ries's theorem, in case the deformations of the environment are characterized by ruptures.

In this paper, this hypothesis is generalized in the case of viscoelastic deformation of anisotropic capillary-porous materials under conditions of temperature-humidity loading, taking into account the phase transition at the boundary of the zone of moisture evaporation. In particular, in the damp zone of the drying process, wood is considered as an orthotropic unsaturated polyphase capillary-porous material with taking into account viscoelastic properties, and in the dry zone, the deformation process is described by the equation of linear viscoelasticity with taking into account the orthotropy of the thermo-mechanical characteristics and the drying up of the material.

## II. STATEMENT OF THE PROBLEM AND THE RESULTS OF THE STUDY

For modeling the rheological behavior of colloidal capillary-porous materials, in particular wood, a model of a heterogeneous system with double porosity for a saturated system is used. Wood is considered as a three-phase system, which consists of wood material (solid phase), liquid and steam-air phases [4]. The distinction of this approach is that the wood is characterized by viscoelastic properties and explicitly describes the volumetric contents of each phase.

Deformation-relaxation processes are described by the following relations:

$$\begin{aligned} \varepsilon_{ij} = K_{ijke}^s (\sigma_{ke} + \bar{\alpha}_{12} \delta_{ke} + \int_0^\tau K_{ijke}^s (\tau - \tau') \cdot \\ \cdot (\sigma_{ij} + \bar{\alpha}_{12} \delta_{ke}) d\tau') + + K_{ijke}^f (\sigma_{ke} + \bar{\beta}_{12} \delta_{ke} + \\ + \int_0^\tau K_{ijke}^f (\tau - \tau') \cdot (\sigma_{ke} + \bar{\beta}_{12} \delta_{ke}) d\tau') \Big) + \alpha_{ij}^s (\Delta U(\tau)), \end{aligned} \quad (1)$$

where  $\varepsilon_{ij}, \sigma_{ij}$  – components of deformations and stresses,  $K_{ijke}^s = K_{ijke}^s(0)/\alpha_3; K_{ijke}^f = K_{ijke}^f(0)/(1-m_\Pi)$ ;  $\bar{\alpha}_{12} = \alpha_1 p_1 + \alpha_2 p_2; \bar{\beta}_{12} = \beta_1 p_1 + \beta_2 p_2$ ;  $\beta_i = m_\Pi \alpha_{ni} + \alpha_{ki}(1-m_\Pi); \alpha_i = \alpha_{ni} m_\Pi + \alpha_{ki} m_k$ ,

$(i=1,2)$ ;  $\alpha_{n1} + \alpha_{n2} = 1$ ;  $\alpha_{k1} + \alpha_{k2} = 1$ ;  $K_{ijke}(0)$  - tensor of instantaneous sensibility,  $\Pi^{-1}$ ;  $\beta_{ij}$  - coefficient of moist expansion;  $K_{ijke}(\tau - \tau')$  - tensor of creep velocity functions;  $\Delta U(\tau) = U(\tau, x) - U_0(\tau)$  - the difference between the current humidity of the wood and its initial value;  $p$  - pressure;  $\tau$  - time;  $m_{II}$  - porosity, which is determined by the ratio of the volume of macropore to the volume of the material;  $m_k$  - porosity, which is determined by the ratio of volume of capillaries to the volume of cell walls;  $\alpha_{ni}, \alpha_{ki} (i=1,2)$  - the contents of the liquid and vapor phases in the volume of pores and capillaries accordingly; the indices  $f$  refer to the effective values (woody skeleton);  $S$  - to the material of wood material;  $II$  - to the macropore system;  $k$  - to the system of capillaries;  $\alpha_1, \alpha_2, \alpha_3$  - volumetric contents of steam-air, liquid and solid phases;  $\delta_{ke}$  - unit tensor.

In the dried area, the rheological behavior of wood is described by linear equations of viscoelasticity with consideration of drying up. They include equilibrium equations  $\partial\sigma_{ij}/\partial x_{ij} = 0$  and linear integral equations of consequence creep for anisotropic environment.

$$\sigma_{ij}(\tau) = \beta_{ij} \left( \Delta U(\tau) + \int_0^\tau K_{ijke}^s(\tau - \tau') d\sigma_{ke}(\tau') \right) \quad (2)$$

Generally, the hypothesis of preserving the residual stresses in the wood as a capillary-porous viscoelastic environment for the phase transition in the case when the material in the damp state during drying process is a polyphase viscoelastic environment, and in the dry state, the wood is described by the equation of consequential creep.

Consider that at the time point  $\tau = \tau^*$  and at the point  $x = \xi(\tau^*)$ , there is a transition from one zone to another. For  $0 < \tau < \tau^*$  the relation between tensions and deformations with material drying up taken into account is rarely (2) described by the equations

$$\begin{aligned} \sigma_{zaa}^{(1)}(\tau^*) = & - \int_0^{\tau^*} R_{ijke}^{(2)s}(\tau^* - \tau') \varepsilon_{ij}^{(1)} d\tau' - \\ & - \beta_{ij}^{(1)} \Delta(U(\tau')). \end{aligned} \quad (3)$$

For the unsaturated wet region, the stress-strain state of wood, taking into account (1), is described by the relations

$$\begin{aligned} \sigma_{ij}^{(2)}(\tau) = & R_{ijke}^{(2)s} \left( \varepsilon_{ke}^{(2)} + \tilde{\alpha}_{12} \delta_{ke} + \int_0^\tau R_{ijke}^{(2)s}(\tau - \tau') \cdot \right. \\ & \cdot \left. (\varepsilon_{ij}^{(2)} + \tilde{\alpha}_{12} \delta_{ke}) d\tau' \right) + R_{ijke}^{(2)f} \left( \varepsilon_{ke}^{(2)} + \tilde{\beta}_{12} \delta_{ke} + \right. \\ & + \int_0^\tau R_{ijke}^{(2)f}(\tau - \tau') \cdot \left. (\varepsilon_{ij}^{(2)} + \tilde{\beta}_{12} \delta_{ij}) d\tau' \right) + \\ & + \tilde{\alpha}_{ij}^{(s)} (\Delta U(\tau)), \end{aligned} \quad (4)$$

where  $R_{ijke}^{(2)}$  - components of the relaxation function tensor, which are determined by components of the creep tensor  $K_{ijke}^{(s)}$ ; coefficients  $\tilde{\alpha}_{12}$ ,  $\tilde{\beta}_{12}$  and  $\tilde{\alpha}_{ij}^{(s)}$  associated with the corresponding dependencies with  $\bar{\alpha}_{12}$ ,  $\bar{\beta}_{12}$  and  $\alpha_{ij}^{(s)}$ .

The components of viscoelastic stresses are obtained by eliminating instantaneous-elastic components. Taking into account that for the phase transition the values of the components of viscoelastic stresses remain the same, we find the components of the wood deformations, assuming that from the very beginning of deformation in the damp state and till the time point  $\tau = \tau^*$  components of the viscoelastic tensions are equal:

$$\begin{aligned} \sigma_{ij}^{(2)}(\tau^*) = & \int_0^{\tau^*} R_{ijke}^{(2)s}(\tau^* - \tau') \cdot (\varepsilon_{ij}^{(2)} + \tilde{\alpha}_{12} \delta_{ke}) d\tau' + \\ & + \int_0^{\tau^*} R_{ijke}^{(2)f}(\tau^* - \tau') \cdot (\varepsilon_{ij}^{(2)} + \tilde{\beta}_{12} \delta_{ij}) d\tau' + \\ & + \tilde{\alpha}_{ij}^{(s)} (\Delta U(\tau^*)). \end{aligned} \quad (5)$$

The ratio is represented as follows:

$$\begin{aligned} \sigma_{ij}^{(2)}(\tau^*) = & \int_0^{\tau^*} \left( R_{ijke}^{(2)s}(\tau^* - \tau') + R_{ijke}^{(2)f}(\tau^* - \tau') \right) \cdot \\ & \cdot \varepsilon_{ij}^{(2)} d\tau' + \int_0^{\tau^*} R_{ijke}^{(2)s}(\tau^* - \tau') \cdot \tilde{\alpha}_{12} \delta_{ke} d\tau' + \\ & + \int_0^{\tau^*} R_{ijke}^{(2)f}(\tau^* - \tau') \cdot \tilde{\beta}_{12} \delta_{ij} d\tau' + \tilde{\alpha}_{ij}^{(s)} \Delta(U(\tau^*)). \end{aligned} \quad (6)$$

Let's bring in the designation

$$\begin{aligned} \int_0^{\tau^*} R_{ijke}^{(2)s}(\tau^* - \tau') \cdot \tilde{\alpha}_{12} \delta_{ke} d\tau' &= R_\alpha(\tau^*), \\ \int_0^{\tau^*} R_{ijke}^{(2)f}(\tau^* - \tau') \cdot \tilde{\beta}_{12} \delta_{ij} d\tau' &= R_\beta(\tau^*) \end{aligned}$$

Therefore, we receive

$$\begin{aligned} \sigma_{ij}^{(2)}(\tau^*) = & \int_0^{\tau^*} R_{ijke}^{(2)s}(\tau^* - \tau') + R_{ijke}^{(2)f}(\tau^* - \tau') \cdot \\ & \cdot \varepsilon_{ij}^{(2)} d\tau' + R_\alpha(\tau^*) + R_\beta(\tau^*) + \tilde{\alpha}_{ij}^{(s)} \Delta(U(\tau^*)). \end{aligned} \quad (7)$$

Let's consider the ratio of the rheological behavior of wood as an unsaturated three-phase environment in the following form

$$\begin{aligned} \sigma_{ij}^{(2)}(\tau) = & R_{ijke}^{(2)sf} + R_{ijke}^{(2)s} \bar{R}_\alpha(\tau') + R_{ijke}^{(2)f} \bar{R}_\beta(\tau') + \\ & + R_{ijke}^{(2)s} \int_0^\tau R_{ijke}^{(2)s}(\tau - \tau') \varepsilon_{ij}^{(2)} d\tau' + \\ & + R_{ijke}^{(2)f} \int_0^\tau R_{ijke}^{(2)f}(\tau - \tau') \varepsilon_{ij}^{(2)} d\tau' + \tilde{\alpha}_{ij}^{(s)} \Delta(U(\tau)), \end{aligned} \quad (8)$$

where

$$\begin{aligned}\bar{R}_\alpha(\tau') &= \int_0^\tau R_{ijke}^{(2)s}(\tau - \tau') \tilde{\alpha}_{ij} \delta_{ke} d\tau'; \\ \bar{R}_\beta(\tau') &= \int_0^\tau R_{ijke}^{(2)f}(\tau - \tau') \tilde{\beta}_{12} \delta_{ij} d\tau'.\end{aligned}$$

Equating the integral expressions for the components of viscoelastic tensions, with (7) and (8) we receive

$$\begin{aligned}\varepsilon_{ij}^{(2)} &= -\frac{R_{ijke}^{(2)s}(\tau^* - \tau') \varepsilon_{ij}^{(1)}(\tau')}{R_{ijke}^{(2)s}(\tau^* - \tau') + R_{ijke}^{(2)f}(\tau^* - \tau')} - \\ &- \frac{R_\alpha(\tau^*) - R_\beta(\tau^*) - \beta_{ij}^{(1)}(\Delta U(\tau')) - \tilde{\alpha}_{ij}^{(s)} \Delta(U(\tau^*))}{R_{ijke}^{(2)s}(\tau^* - \tau') + R_{ijke}^{(2)f}(\tau^* - \tau')}\end{aligned}\quad (9)$$

Taking into account that during the process of phase transition, components of viscoelastic stresses are retained, taking into account (4), (8), (9), we write down the defining relations for viscoelastic deformation of wood

$$\begin{aligned}\sigma_{ij}^{(2)}(\tau) &= R_{ijke}^{(2)s}(\varepsilon_{ke}^{(2)} + \tilde{\alpha}_{12} \delta_{ke}) + R_{ijke}^{(2)f} \cdot \\ &\cdot (\varepsilon_{ke}^{(2)} + \tilde{\beta}_{12} \delta_{ke}) + R_{ijke}^{(2)s} \int_0^\tau R_{ijke}^{(2)s}(\tau - \tau') \tilde{\alpha}_{12} \delta_{ke} d\tau' + \\ &+ R_{ij}^{(2)f} \int_0^\tau R_{ijke}^{(2)f}(\tau - \tau') \tilde{\beta}_{12} \delta_{ke} d\tau' + \\ &+ \int_0^* \frac{R_{ijke}^{(2)s} R_{ijke}^{(2)s}(\tau - \tau') + R_{ijke}^{(2)f} R_{ijke}^{(2)f}(\tau - \tau')}{R_{ijke}^{(2)s}(\tau^* - \tau') + R_{ijke}^{(2)f}(\tau^* - \tau')} \cdot \\ &\cdot (R_{ijke}^{(1)}(\tau^* - \tau') d\tau' - \tilde{\alpha}_{ij}^{(s)} \Delta(U(\tau^*))) d\tau' + \\ &+ (R_\alpha(\tau^*) - R_\beta(\tau^*) - \beta_{ij}^{(1)}(\Delta U(\tau')) + \\ &+ R_{ijke}^{(2)s} \int_{\tau^*}^\tau R_{ijke}^{(2)s}(\tau - \tau') \varepsilon_{ij}^{(2)} d\tau' + R_{ij}^{(2)f} \cdot \\ &\cdot \int_{\tau^*}^\tau R_{ijke}^{(2)f}(\tau - \tau') \varepsilon_{ij}^{(2)} d\tau' + \tilde{\alpha}_{ij}^{(s)} \Delta(U(\tau))).\end{aligned}\quad (10)$$

Thus, the obtained relations take into account the deformation-relaxation processes of the wood as a multiphase unsaturated environment for both before and after the evaporation zone of the moisture. In particular, the effect on stress relaxation in the wood of the prehistory of deformation to the phase transition is taken into account.

### III. RESULTS OF THE NUMERICAL EXPERIMENT

The process of deformation of capillary-porous materials is characterized by a change of their volume. This causes a change of phase environment's size, which significantly impedes researching the rheological behavior of the material. Therefore, we consider the contribution of residual stresses to relaxation for a one-dimensional viscosity case taking into account the above-described phase transition. In this case, the functions of the rheological behavior of the wood taking into account the mechanism of accumulation of residual

deformations [5,6]  $K^{(s)}(\tau - \tau')$  and  $K^{(f)}(\tau - \tau')$  we choose in the form

$$\begin{aligned}R^{(2)(i)}(\tau - \tau') &= \left[ a_0 - \sum_{i=1}^M a_i \exp(-b_i \tau) \right] h(\tau) \cdot \\ &\cdot h(\tau_0 - \tau) - \left[ a_0 - \sum_{i=1}^M \alpha_i \exp(-\beta_i(\tau - \tau_0)) \right] \cdot \\ &\cdot h(\tau - \tau_0),\end{aligned}\quad (11)$$

where  $h(\tau)$  Heaviside's function, and the unknown coefficients  $a_i, b_i, \alpha_i, \beta_i$  are determined by the method of least squares based on the approximation of experimental data on the creep of samples of timber under load and after unloading. In the general case, they are functions of temperature  $T(x, \tau)$  and moisture  $U(x, \tau)$ . To do this we use the method of minimum squares [7]. To quantify the difference, a statistical criterion was used based on the correlation coefficients [7].

To determine the parameters  $\alpha_{i0}$  ( $i = 1, 2, 3$ ), taking into account the change in humidity, the correlation are obtained taking into account the conditions of additivity and the uniform distribution of phases over the wood regions. Then, in accordance with (9), (10), (11), (12) we obtain relations for determining the effect of residual stresses on wood during the phase transition for a one-dimensional deformation case

$$\begin{aligned}\Delta \sigma(\tau^*) &= \int_0^* (\bar{E} \exp((\tau^* - \tau') / \tau_{per}(U, T)) \cdot \\ &\cdot (\bar{R}_0^{(2)s} R^{(2)s}(\tau - \tau') + \bar{R}_0^{(2)f} R^{(2)f}(\tau - \tau') + \\ &+ \bar{R}^{(2)f}(\tau - \tau')) / (\bar{R}_0^{(2)s} R^{(2)s}(\tau^* - \tau') + \\ &+ \bar{R}_0^{(2)f} R^{(2)f}(\tau^* - \tau') + \bar{R}^{(2)f}(\tau^* - \tau')) \cdot \\ &\cdot \varepsilon^{(1)}(\tau') d\tau',\end{aligned}\quad (12)$$

where

$$\begin{aligned}\bar{R}_0^{(2)s} &= R_0^{(2)s} (1 - \varepsilon) \alpha_{20} \rho_2 / \alpha_{30}; \\ \bar{R}_0^{(2)f} &= R_0^{(2)f} (1 - \varepsilon) (\alpha_{30} + b(\alpha_{20}); \\ \bar{R}^{(2)f} &= \int_0^* R_0^{(2)f} R^{(2)f}(\tau^* - \tau) (\alpha_{30} + \alpha_{20}) \cdot \\ &\cdot \frac{\rho}{(\alpha_{30} + b\alpha_{20})} d\tau^*; \\ \alpha_{10} &= 1 - \rho_w \left( \frac{1}{\rho_3} + \frac{U}{100\rho_1} \right) \cdot \frac{100}{100 + U}; \\ \alpha_{20} &= \frac{1}{\rho_1 - \rho_2} \left( \rho_w \left( \frac{1}{\rho_3} + \frac{U}{100\rho_2} \right) \cdot \right. \\ &\left. \cdot \frac{100}{100 + U} - \rho_1 \right);\end{aligned}$$

$$\alpha_{30} = \frac{1}{\rho_1 - \rho_2} (\rho_w (\rho_1 - \rho_3 - 1) \cdot \left( 1 + \frac{1}{\rho_3} \frac{U}{100\rho_2} \right) \cdot \frac{100}{100+U} - \rho_1;$$

$$\rho_w = \begin{cases} k_{\alpha 1} \rho_{12} \frac{100+U}{100+k_{\alpha 2} U}, & U \leq 30\%; \\ k_{\alpha 3} \rho_{12} (1+0,1U), & U > 30\%, \end{cases}$$

where  $k_{\alpha 1}, k_{\alpha 3}, \rho_{12}, b$  – coefficients determined by the properties of wood [5].

For the case of deformation changes at the moment of the phase transition, which are characterized by a change in density, depending on the change in humidity. Also, the linear dependence of modulus of elasticity on changes in humidity is taken.

The nature of the distribution of stress relaxation curves shows that the consideration of viscoelastic stresses during the phase transition in the wood during the drying process differs from stress relaxation curves without taking into account residual viscoelastic deformations.

In fig. 1 and 2 graphic dependences of relaxation of viscoelastic stresses in wood with a base density  $\rho=530$  kg/m<sup>3</sup> are given.

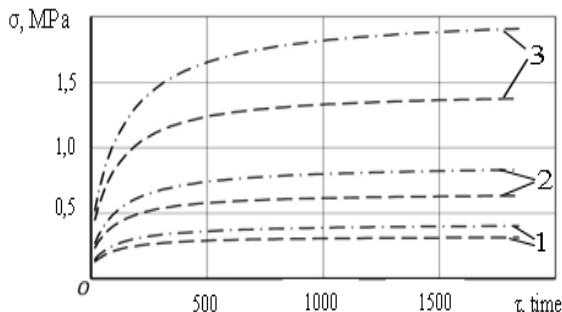


Fig.1 Dependencies of relaxation of stresses for materials of wood in the radial direction  $\rho_0=550$  (— · —),  $\rho_0=400$  (— · —) for different values of initial humidity (1–10%, 2–15%, 3–20%).

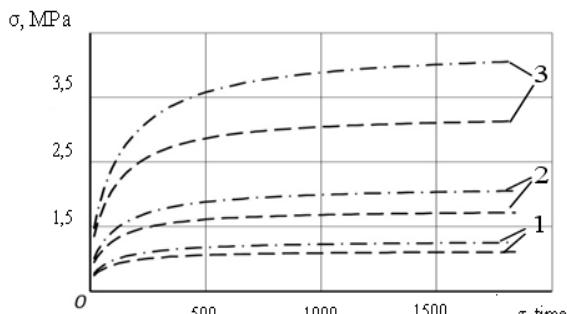


Fig.2 Dependencies of relaxation of stresses for materials of wood in the tangential direction  $\rho_0=550$  (— · —),  $\rho_0=400$  (— · —) for different values of initial humidity (1–10%, 2–15%, 3–20%).

#### IV. CONCLUSION

The mathematical model of determination of viscoelastic deformation of capillary-porous materials as a three-phase system with including anisotropy of thermo mechanical characteristics is given.

The regularities of the influence of transfer mechanisms on processes of viscoelastic deformation in the solid, liquid and vapor phases for wood are established.

Applied software for numerical implementation of mathematical models is developed.

A generalization of the hypothesis of the saving of irreversible deformations in the case of viscoelastic deformation of capillary-porous materials, taking into account the phase transition at the boundary of the evaporation of moisture is obtained.

#### REFERENCES

- [1] J. Bodic, A. Jayne, "Mechanics of Wood and Composites," *Van Nostrand Reinhold*, New York, 1982, pp. 712.
- [2] M. S. Mozharovskyy, "The theory of elasticity, plasticity and creep," *High school*, Kyiv, 2002, p. 313.
- [3] M. M. Dubina, B. A. Krasovitskyy, A. S. Lozovskiy, F. S. Popov, "The thermal and mechanical interactions of engineering structures with frozen soils," *Science*, Novosibirsk, 1977, p. 144.
- [4] O. R. Dronyak "The mathematical modeling, computer optimization of technologies, equipment parameters and the systems of the forest complex," *Interuniversity collection of scientific works, Voronezh State Forestry Academy*, Voronezh, 2001, pp. 132-139.
- [5] B. N. Ugolyev, "The wood science with the basics of forest commodity science," *the textbook for the forestry universities, State Forest University*, Moskow, 2002, p. 340.
- [6] Ya. I. Sokolovskyy, Y. V. Andrashek, "Methodology and results of experimental researches of wood rheological behavior," *Scientific periodical of the Ukrainian State Forestry University of Ukraine, Collection of Scientific and Technical Works*, Lviv, 1999, pp. 15-26.
- [7] A. K. Malmeister, G. A. Tamuzh, "Resistance of polymer and composite materials," *Temers-Riga, Zinatie*, 1980, p. 572

# Web Application for Air Quality Monitoring

Yevhen Kedrin<sup>1</sup>, Iryna Voytyuk<sup>2</sup>, Serhii Tryshkaliuk<sup>3</sup>, Mykhailo Shpintal<sup>4</sup>

Department of Computer Science, Ternopil National Economic University, UKRAINE, Ternopil, 8 A. Chehova str., email: [place.ev@gmail.com](mailto:place.ev@gmail.com)<sup>1</sup>, [i.voytyuk@tneu.edu.ua](mailto:i.voytyuk@tneu.edu.ua)<sup>2</sup>, [grisha.semyonav@gmail.com](mailto:grisha.semyonav@gmail.com)<sup>3</sup>, [m.shpintal@tneu.edu.ua](mailto:m.shpintal@tneu.edu.ua)<sup>4</sup>

**Abstract:** Web application for air quality control as part of ecological monitoring system presented in this paper. It is proposed the database architecture and modules for data processing and visualising. Proposed application enables to monitor air pollutions from motor transport not only in measurement points, but in different points of city using mathematical models.

**Keywords:** monitoring system, air quality control, vehicular pollution, web application.

## I. INTRODUCTION

One of the biggest environmental pollutants in large cities is motor transport. During the fuel combustion the emissions such as sulfur dioxide ( $\text{SO}_2$ ), particulate matter ( $\text{PM}_{10}$ ), fine particulate matter ( $\text{PM}_{2.5}$ ), nitrogen dioxide ( $\text{NO}_2$ ), carbon monoxide ( $\text{CO}$ ), ozone ( $\text{O}_3$ ) and others are polluting the air. For pollution assessing the measurement scale is using based on Air Quality Index (AQI), which determine whether a harmful substances are under permissible level. For example, the lowest level of harmful substances called "Very low", next are "Low", "Medium", "High", "Very high". Different countries have their own measurement scales. The Common Air Quality Index (CAQI) is an air quality index used in Europe since 2006. In November 2017, the European Environment Agency announced the European Air Quality Index (EAQI) and started encouraging its use on websites and for other ways of informing the public about air quality. [1] The sample of CAQI are given in Table 1.

TABLE 1. COMMON AIR QUALITY INDEX (CAQI)

| Qualitative name | Index or sub-index | Pollutant (hourly) density in $\mu\text{g}/\text{m}^3$ |                  |              |                   |
|------------------|--------------------|--|------------------|--------------|-------------------|
|                  |                    | $\text{NO}_2$  | $\text{PM}_{10}$ | $\text{O}_3$ | $\text{PM}_{2.5}$ |
| Very low         | 0–25               | 0–50   | 0–25             | 0–60         | 0–15              |
| Low              | 25–50              | 50–100   | 25–50            | 60–120       | 15–30             |
| Medium           | 50–75              | 100–200  | 50–90            | 120–180      | 30–55             |
| High             | 75–100             | 200–400  | 90–180           | 180–240      | 55–110            |
| Very high        | >100               | >400   | >180             | >240         | >110              |

The level of these pollutants in the air affects the well-being and health of the city's population. At high levels of pollution, it is recommended to reduce or completely abandon the activity and exercises in the open air. This is especially true for children, the elderly and people with heart or respiratory diseases. In this regard, the population should be provided with quick and easy access to relevant data on the air pollution in the city regions.

## II. TASK STATEMENT

To date, there are organizations in Ukraine that have means

for periodic measurement of harmful emissions from vehicles. However, such approach does not provide an operational monitoring of the dynamics of these emissions. In the field of mathematical modeling, a series of diffusion mathematical models have been developed that describe the processes of harmful emissions diffusion into the atmosphere [2-5]. At the same time, the results of the application of these models are poorly correlated with actual measurement data. As a result, it is impossible to detect and react in a timely manner to raising the pollution level.

It is necessary to use modern technologies and devices that will make monitoring of air pollution in the cities more accurate and continuous. Establish on the city territory of the sensors for measurement different substances and control of traffic flows. In order for the whole area of the city to be covered, not only certain areas, it is necessary to create mathematical models and software that will show the spatial distribution of harmful substances. Access to all data should be provided via the Internet. Every resident should be able to find out what the current state of the air in the city is. Viewing the measurements should be simple, understandable, using conditional labels and images on the map. It should be possible to know the measurements for certain periods or only for certain parameters. Additional opportunities are also needed, such as informing the public, publishing news, announcements, photo materials, conducting discussions, exchanging ideas and suggestions. For a better understanding of the problem, it is necessary analyzing already existing systems of the same purpose and creating its own, which will be useful both for ordinary residents of the city and for specialized workers in the field of ecology. This will be a good step towards improving the environment and human health. The purpose of research is to plan the overall structure of the system for improving the quality and speed of air pollution monitoring by road, and to facilitate the obtaining of this information for the population of cities through the web-resource.

## III. SYSTEM REQUIREMENTS AND ARCHITECTURE DESIGN

In order to automate the air pollution monitoring, it was decided to create a system consisting of the three components described below.

1) Hardware including sensors that measure the different parameters, such as pollutant concentration, traffic flow, humidity, temperature. They will be located in different parts of the city, mainly in central parts and in places where there is a large crowd of vehicles, such as a crossroads. Data from such sensors will be transmitted using mobile communication - GSM-modules. To ensure the proper level of storage, all devices will be placed in specially designed containers or in

existing ones, for example, in places where traffic control cameras are located, and must be agreed with local authorities.

2) Software, based on mathematical models, calculations and measurements from sensors, will simulate the spatial distribution of pollution on a city map. This will allow to see the results of measurements not only in the places where they are conducted, but also to predict the level of pollution at each point of the map. The papers [6,7] show the method of solving the task of identifying mathematical models for environmental monitoring of air pollution by motor vehicles based on the use of fireflies algorithm or the bee colony algorithm. It is this research that underpins the development of a mathematical model that will be used in our system.

3) Web application that is designed to conveniently visualize the resulting measurements.

The web application must flexibly interact with other parts of the system and with any other sensors or datasets that will be available for use. The air quality level can be depicted on a city map in the form of points where the sensors are located, or as a result of mathematical modelling the spatial distribution of pollution.

As a result of the analysis of existing alternative solutions the number of functional and non-functional system requirements were listed, namely: possibility of showing the sensor measurements and simulation results on city map; ability to choose a period for showing measurements; ability to view detailed information at the selected point of the map; ability to take information through the API from various sources: sensors, other software systems, Excel files; working with different indicators and possibility of combining them on the map; development of administrative profile with ability to adjust the map showing and indicators, edit news and other materials; development of news page, including the latest news on the main page; possibility to place announcements on the main page; development of the blog; possibility to leave comments under every material; showing the contact data and feedback form; development of the forum; development of the photo gallery; provide a system design that will allow viewing site data in various browsers and devices (PC, tablet, phone).

Taking into account defined requirements to the system, a database was designed. The structure and relations of tables for working with maps and measurements showed in Fig. 1.

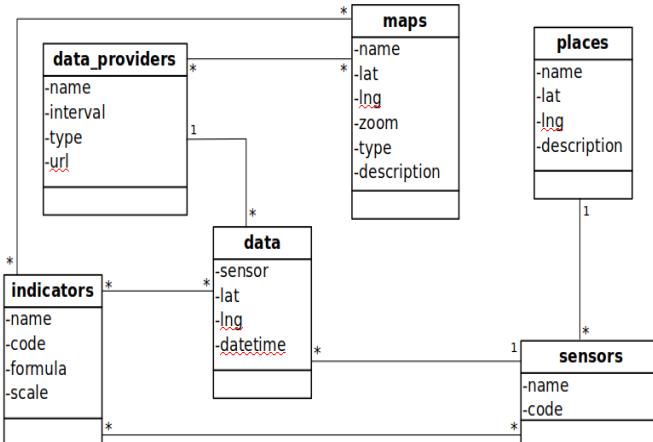


Fig. 1. Database structure

#### IV. WEB APPLICATION IMPLEMENTATION

Web application as one of the components of the system for monitoring atmospheric pollution from vehicles, is designed to visualize the results of measurement and modeling the spatial distribution of pollution (fields of concentration of harmful emissions). As a result of research, a web-based system project was developed and implemented in PHP programming language. As a database, MySQL was used. CSS Media Queries, Bootstrap, were used to design a convenient and compatible site with various devices. For fast work due to AJAX interaction and dynamic interfaces, the JavaScript library for jQuery is used.

As geographic maps the Google Maps was used with additional data such as the traffic flow rate. The Google Maps APIs [8] allow to flexibly customize viewing and map interaction, add graphic objects, and various data types that are well suited for pollutions visualization. There are currently two maps on the website. The first depicts the points with the actual measurements coming from the sensors (Fig. 2). Each of the points has own color and number corresponding to the AQI scale, which is under the map.

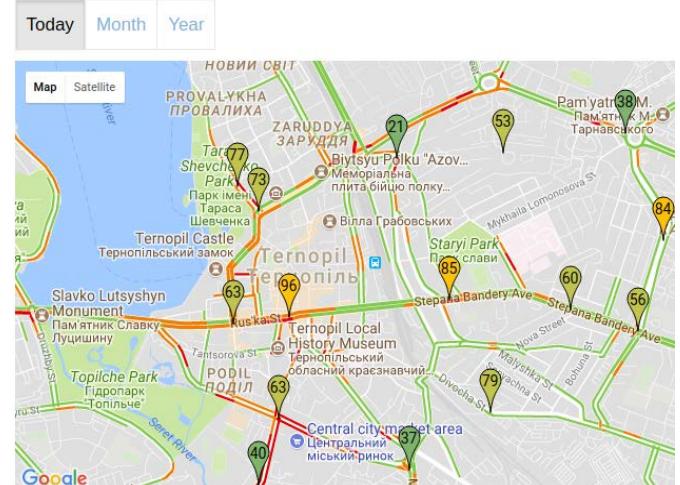


Fig. 2. Map with sensor indicators

On the second map are the simulated results calculated on mathematical model, in which one can observe the spatial distribution of pollution (Fig. 3).

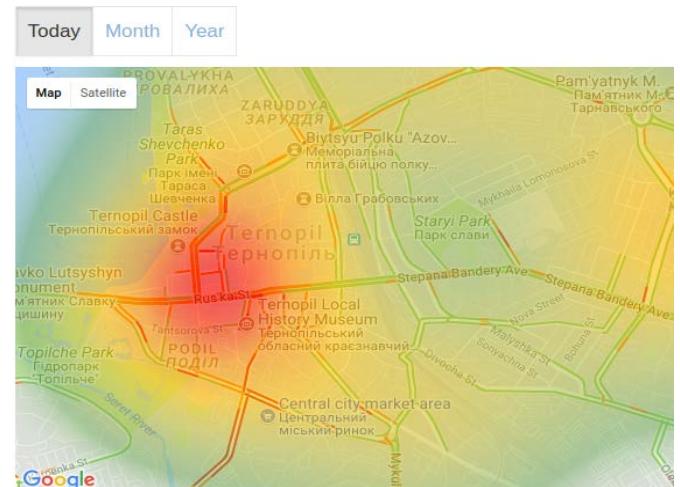


Fig. 3. Map with spatial distribution

There is at the top of each map a switch where it is possible to specify the period for which measure to be displayed: today, month, or year. Also, site users can view detailed information by clicking on a specific point on the map. Fig. 4 shows an example of an auxiliary information window, which, in addition to the indicators shown on the map, also contains additional ones such as humidity and air temperature. The content and method of displaying items in this window are configured from the administrative part.

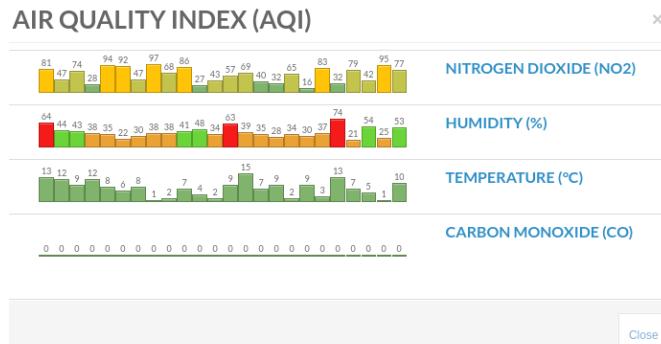


Fig. 4. Information screen

In addition to viewing maps, the website has pages with a wealth of useful information, news and blogs with possibility to add the interesting articles. In order for site users to ask questions or to offer an idea, there is a page with contact information and a feedback form. Also on the site is a forum that allows to discuss topics related to air pollution and how to improve the situation with the environment.

The administrative profile allows to create new admins and give them only certain permissions. It is possible to conveniently edit the content of the site, create news, write articles. The specialized section of the administrative profile is the control of the display of information on air pollution.

The system is called "Our city" and contains the following administrative sections: maps, providers, indicators, placement, sensors. The main menu view is shown in Fig. 5.

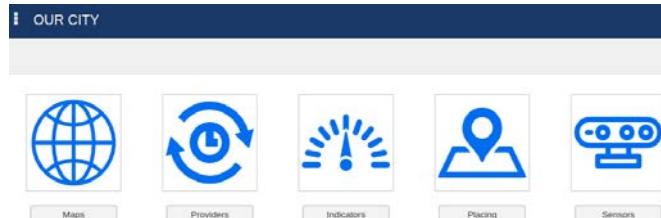


Fig. 5. Administrative menu

"Maps". In this section the list of all geographic maps that will be available for viewing on the site is shown. When creating a new map, it can be flexibly customized depending on the needs and the data it should display. So, when creating it is necessary to specify the name, latitude and longitude, the level of convergence, the type of relief, if necessary, a description (Fig. 6). Also it is necessary to specify one or more data providers to choose which data to display.

Each provider can provide a large amount of information and indicators, but which of them should be displayed, in what form, as well as where - on the map or in the detailed view window at the selected point - it is configured in the Indicators sub-tab (Fig. 7). Here can be selected any indicator

and specify its appearance on the map (distribution chart or point of a certain color with the number), and a detailed view (diagram, graph, etc).

This screenshot shows a configuration form for a map entry. It includes fields for 'Name' (Ternopil - Math forecasts), 'Latitude' (49.551392), 'Longitude' (25.596640), 'Zoom' (14), 'Type' (Roadmap), 'Providers' (FCIT - Pollution monitoring - Math model, Model - Sample data), 'Status' (Published), and a rich text editor for 'Description'.

Fig. 6. Form for map editing

This screenshot shows a configuration form for map indicators. It features a table with columns for 'Indicator' (Nitrogen dioxide (NO2), Humidity (%), Temperature (°C), Carbon monoxide (CO)), 'Map view' (Headmap, Chart), and 'Popup view' (Chart). The 'Map view' dropdown for NO2 is set to 'Headmap' and 'Popup view' to 'Chart'. The other rows have 'Chart' selected for both map and popup views.

Fig. 7. Form for map indicators editing

"Providers". This section contains a list of all data sources from which the site may receive information. Connection can take place via HTTP or FTP protocol. Thus, it is possible to obtain data not only from other parts of the monitoring system - sensors and programs with implemented mathematical models, but also to third-party companies that carry their own measurements or have their equipment to sanitary stations. Also, here it is indicated, with what interval should apply for the updated data.

“Indicators”. This section contains a list of all parameters, and metrics that the system works with. For each indicator, it is possible to specify the name, code and scale of measurements, where for each level is given the color and numbers min and max (Fig. 8). It is also possible to specify a formula for converting values or complex calculations. For example, one indicator shows actual values from the sensor, named NO<sub>2</sub>, and an additional indicator, named NO<sub>2</sub>-AQI, will display measurements of nitrogen dioxide translated into the air quality index. Or it was created a common indicator AQI, which in its calculations will use several different sensors measurement.

| Scale | Min | Max | Color   |
|-------|-----|-----|---------|
|       | 0   | 40  | #80b46c |
|       | 41  | 80  | #c3c44d |
|       | 81  | 180 | #ffc308 |
|       | 181 | 280 | #ff9300 |
|       | 281 | 400 | #ff0000 |
|       | 400 | 999 | #a82b22 |

Fig. 8. Form for indicator editing

“Placing”. This list lists all the locations in which the system receives measurements. When editing it is necessary to specify a name, latitude, longitude, description.

This section is intended for grouping sensors, because there can be a lot of them in one place.

“Sensors”. In this section it can view the list of all sensors that the system uses. When connecting a new sensor, it need to add it to this list by specifying its name, code, list of measurements it passes on (indicators) and indicate its location. With this information the system will automatically detect sensor or communication problems, responding to the lack of measurements or their inaccuracy.

This structure makes the system flexible, allowing it to create new maps without software changes, to connect different data sources, to work with various indicators, even if in the future it will be necessary to measure, say, water or soil. Also, there is no hard bind to other components of the system, the application will be able to work correctly even at the time of their modification, and in turn will not affect the work of other components, such as sensors, when work on the site itself.

## V. CONCLUSION

The processes of air pollution by motor transport are one of the most important problems of large cities. To solve this problem it is necessary to ensure operational monitoring of the dynamics of harmful emissions. The development of new models based on real monitoring data will increase the accuracy of both reflection and prediction of the dynamics of atmospheric pollution, depending on the intensity of traffic flows, and the creation of an appropriate software system for their implementation, as well as a web resource to visualizing the current situation of atmospheric pollution, will serve as an active tool for municipal services that deal with traffic flow management, environmental protection, and for the city's population.

Within this research the general structure of the system was described and a web application was developed.

## ACKNOWLEDGMENT

This research has been supported by National Grant of Ministry of Education and Science of Ukraine "Mathematical tools and software for control the air pollution from vehicles" (0116U005507).

## REFERENCES

- [1] "Impact assessment on the Air Quality Package (summary)", SWD/2013/0532 final, EC, 2013.
- [2] M. A. Sutton, C. M. Howard, J. W. Erisman. *The European Nitrogen Assessment: Sources, Effects and Policy Perspectives*, Cambridge University Press, 2011.
- [3] Outdoor air pollution a leading environmental cause of cancer deaths, Press Release No 221, International Agency for Research on Cancer, World Health Organization, Lyon, France, 17 October 2013.
- [4] "Global Road Safety Facility", The World Bank and Institute for Health Metrics and Evaluation, Transport for Health: The Global Burden of Disease From Motorized Road Transport, IHME; the World Bank, Seattle, WA; Washington, DC, 2014.
- [5] "Regulation (EC) No 443/2009 of the European Parliament and of the Council of 23 April 2009 setting emission performance standards for new passenger cars as part of the Community's integrated approach to reduce CO<sub>2</sub> emissions from light-duty vehicles", OJ L 140, EU, 2009, pp. 1–15.
- [6] I. Voytyuk, N. Porplytsya, A. Pukas, T. Dyvak. "Identification the interval difference operators based on artificial bee colony algorithm in task of modeling the air pollution from vehicular traffic", in *Proceedings of 14th International Conference: The Experience of Designing and Application of CAD Systems in Microelectronics*, 2017, pp. 58-62.
- [7] I. F. Voytyuk, S. R. Tryshkaliuk, Ye. S. Kedrin. "Investigation of fireflies algorithm for identification of mathematical model in the task of monitoring pollution by motor transport," in *Proceedings of Conference: Advanced Computer Information Technologies*, Ternopil, 2017, pp. 235-238.
- [8] Google Maps APIs. Available: <https://developers.google.com/maps/web/>

# Computer Network Technologies

# Chain of Clusters for Improving Network Lifetime of Sensor Network

Rashmi L. Jain<sup>1</sup>, Sachin Jain<sup>2</sup>

1. Computer Science and Engineering Department, Rajiv Gandhi College of Engineering & Research, Nagpur, India.  
E-mail: rashmilalitjain@gmail.com

2. Information Technology Department, Yeshwantrao Chavan College of Engineering, Nagpur, India.  
E-mail: sachinjain98440@rediffmail.com

**Abstract:** In wireless sensor network increasing life time of sensor node and there by network is the main motive for development of a protocol for the sensor network. LEACH (Low Energy Adaptive Clustering Hierarchy) is an energy-efficient hierarchical protocol that balances the energy consumption, saves the node energy as compared to flat communication protocols and hence prolongs the lifetime of the network. Here, we planned a new hierarchical cluster based protocol for varied sensor networks. Instead of selecting the cluster head randomly, we include chain forming concept of PEGASIS (Power Efficient Gathering in Sensor Information System). In our work we consider heterogeneous nodes. Unlike LEACH, proposed protocol uses the selection criteria for Cluster-Head depending on the residual energy of the nodes and relative distance of cluster heads. A chain of Cluster Head will be formed using chain formation technique of PEGASIS protocol. Successful implementation of data aggregation has reduced the energy-consumption.

**Keywords:** Energy efficiency, chain formation, LEACH protocol, PEGASIS protocol, Heterogeneous wireless network, Data aggregation.

## I. INTRODUCTION

Wireless sensor networks (WSNs), are generally used for monitoring some phenomena or certain parameters. They are generally formed with ample number of small, resource constrained, sensing nodes which are spread either with some criteria or randomly in a geographical area of interest. Sensing data and communicating them to the data collecting sink are major operations taking place. But major component for success of the wireless sensor network applications is the way of communication among the sensor nodes. This is because more energy of the energy restricted nodes gets drained for communication in comparison with energy consumed in computation or processing the information collected. So for designing any wireless sensing application one has to keep in mind the critical parameter of energy utilization in the network. Delivery of data in a specific time slot is again very essential for the successful task completion of nodes that's why routing parameters are essential. Routing path of the data in WSN determines how data hops from node to node in order to reach the destination when the destination is not directly reachable from the source node. If source and destination nodes are at a shorter distance ( $d < d_0$ , where  $d_0$  is

a cross over distance) then communication energy required is less and is in proportion to  $d^2$ , where as if the distance is more than  $d_0$  then more energy is consumed and is in proportion to  $d^4$ . So if long distance communication is done frequently node battery will drain quickly. Efforts are being taken to reduce this communication power and boost the life of the nodes in the network so that they can serve the purpose for longer time. Such energy efficient wireless sensor network applications are highly demanded, for different scenarios.

The forthcoming paper is divided in following sections. Part 2 will discuss about LEACH protocol and PEGASIS protocols. In section 3, discussion about proposed protocol and radio energy model is done. Section 4 will tell about results of proposed method in comparison with LEACH and PEGASIS protocol. Lastly in section 5 conclusion and future scope of proposed method are discussed.

## II. RELATED WORKS

Protocols have been developed for communication purpose in WSN since its inception. Earlier Direct Transmission, Minimum Energy Transmission (MTE) protocols were used, which are considered as flat routing. W.R.Heinzelman then added a milestone to these protocols with design of LEACH protocol [1]. In the cluster based approach given by LEACH, considerable energy saving was achieved when the base station is distant as compared to the sensor nodes deployment. Lot of work is done on this protocol and various enhancements are available. Chandrashekharan provided another concept of chain based communication, which outperformed as compared to LEACH. But certain limitations were there. In this section we discuss about these protocols in brief.

### A. The LEACH protocol

LEACH protocol is proposed by W.R. Heinzelman [1]. LEACH i.e. Low Energy Adaptive Cluster Hierarchy, is hierarchical clustering protocol. It is self organized protocol and nodes are divided in clusters for sharing information to base station via cluster head (CH). The idea of hierarchical routing approach provided in LEACH is an inspiration, an anchor for development of many other hierarchical protocols, although some of them are having novice idea and are developed independently [2]. This protocol assumes that all the nodes have equal initial energy,  $E_0$ , they are homogeneous in nature and they are randomly placed in the

sensor field. It includes two operations performed in each round of working, namely setup and steady state operation. Clusters of nodes are formed in first phase- setup, where, a part of nodes, P, which is predetermined, elect themselves as CHs as follows.

Here, a random number r, between 0 and 1 is selected by the sensor node. T(n), a threshold value, is considered as given in below mentioned Eq.1. If r is less than T(n), then this node declares itself as CH for present round. The threshold value depends upon the expected percentage P, for number of CH in present round. Only that set of nodes can become a CH that has not yet become CH in the previous (1/P) rounds. Such nodes which are involved in cluster head selection [14], are denoted by set G. This T(n) is calculated by:

$$T(n) = P/(1-P(r \bmod (1/P))), \text{ if } n \in G \quad (1)$$

In the current epoch T(n) value is zero for the nodes who have been the cluster heads in subsequent rounds. Each CH elected transmits an announcement message to other nodes in the network about their role as the new cluster-heads. Other Non-CH nodes select the nearest CH and join that respective cluster as a member. CHs create and broadcast a TDMA plan for their associate nodes for data transmission so that collisions are avoided.

During the steady state phase actual work of the sensor nodes begins. They sense and transmit the data to the cluster-heads. The cluster head receives the data from members and aggregates it to reduce the size, before sending it to the sink and then transmits this data. This data compaction will reduce the communication energy needed during transmission.

In the next round again the setup and steady state operations are repeated.

#### B. The PEGASIS protocol

PEGASIS i.e. Power-Efficient Gathering in Sensor Information Systems (PEGASIS) protocol is proposed by S. Lindsey and C. S. Raghavendra [2]. This is chain based protocol in which each node in network will form a chain and can only communicate with their adjacent node. Each node finds its closest node with the help of signal strength received.

In this, main communicating node, the leader, is chosen based on energy remained with the nodes. Leader will be responsible for gathering data from each node in chain and then sending that data to BS. PEGASIS does not form cluster and uses only leader node for communicating with sink. This will reduce overhead of cluster head selection as well as the bandwidth needed in messaging.

But drawback with LEACH and PEGASIS is that, In LEACH there is no certainty about cluster head selection since CH is selected randomly. In PEGASIS, it may be possible that leader that is responsible for sending network information to BS, will be located at some point which far from BS. So, leader may consume more energy for sending information from longer distance to BS. These two major drawbacks of LEACH and PEGASIS are overcome in our proposed method.

### III. THE PROPOSED PROTOCOL

The proposed protocol uses the advantages of the clustering mechanism of improved LEACH protocol and chain formation technique of PEGASIS protocol. The protocol results in improved life of sensor node which ultimately increases lifetime of network. In this section we will discuss the details of our protocol. Unlike LEACH, proposed protocol uses heterogeneous nodes in the network; also cluster heads are selected on basis of residual energy of nodes. In basic LEACH, selection of CH is depends upon the probability function but this criteria is useful only if energy of nodes in the network is uniform. Second most important issue in LEACH is that, there is no certainty about whether the CHs are distributed uniformly in network or not. This problem of LEACH is resolved by forming clusters according to the distances of nodes along the x-axis.

Secondly, in PEGASIS all nodes have probability to become leader node for a particular round. Therefore there are chances that node selected as leader node for that round may not have sufficient amount of energy to survive in the network for longer time. If the leader node dies then no communication can be done with the base station and results in breakage of network. Also there are chances that selected leader node is located far away from a base station. Therefore, time and energy dissipation for sending network data to base station is high. This issue of PEGASIS protocol is resolved by selecting a leader node which will have maximum amount of energy and which is located nearer to the base station. Only this last leader node will be communicating with the base station.

#### A. Working of protocol

The heterogeneous nodes are deployed randomly in the sensor field and groups are formed by arranging sensor nodes according to x-axis at some fixed point in the network area. This process leads to the formation of clusters in network along x- direction. After creating clusters, energy of each sensor node of cluster is calculated and two nodes with highest energy are selected. First node with the highest energy is the CH node and second highest energy node is stand by CH. If first cluster head dies then sensor node with second highest energy i.e. stand by CH will act as CH. Once this clustering is over, formation of chain between CHs of neighbouring clusters is done. All the sensor nodes of cluster will send their information to the respective CH of the cluster. Then selected CH will perform aggregation of this information and forward this to the CH of neighbouring cluster. CH of neighbouring cluster will add this received information with its own cluster's information and then again it will carry forward this whole information to the next neighbouring CH of cluster. This process will continue till information of every cluster will reaches to the CH of last cluster of the network. Now, this CH will aggregate this whole information in small but meaningful message and send it to the base station. CH of last cluster is selected so that distance between this CH and base station is minimum. Following Fig.1 shows the entire process of clustering and chain formation between CHs.

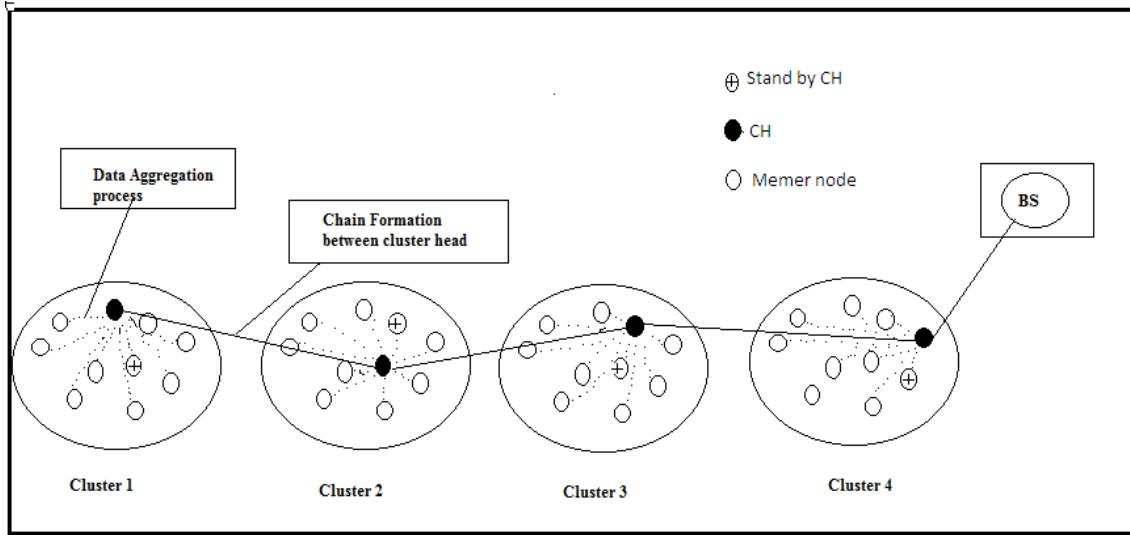


Fig. 1. Cluster and chain formation of nodes.

### B. Radio and energy model

In the proposed method the nodes are communicating to their neighbouring nodes, so long distance communications are avoided. Radio models for a distance less than cross-over distance and for distance greater than cross-over distance are projected by W.R. Heinzelman [1], that are used here.

In our protocol  $d^2$  path loss model is utilized for data transmission, since the nodes communicate the data to their CH in close vicinity. Transmission and reception costs for message with  $k$  bits over a distance  $d$  are as given in Eq. 2 and Eq. 3.

#### Cost for Transmission:

$$ET_x(k,d) = E_{elec} * k + E_{amp} * k * d^2, \quad \text{if } d < d_0 \quad (2)$$

#### Cost for Reception:

$$ER_x(k) = E_{elec} * k \quad (3)$$

In the equations,  $d$  is path-loss exponent depending on distance between source and destination. Here it has been taken of order  $d^2$  for experimentation. Again  $k$ , the number of bits transmitted and received should be less because high amount of energy is spent in transmission and reception of data in comparison data processing. Long distance communication may take place only between the leaders and sink node if it is placed beyond the cross-over limit in the network. The network scenario used is given in the next section.

### C. Network parameters

The proposed protocol has been simulated in one of the most popular and appropriate Network simulator i.e. NS2, version ns2.35. In the scenario considered there are total 40 nodes in network which are placed randomly in the area 1300 X 1000 meters. Each node has different initial energy as the nodes are of heterogeneous type. Initial energies of the nodes are also assigned randomly, indicating values between 0 and 1000. Energy dissipation during data communication is dependent on distance between the source and sink node.

Therefore the energies of the nodes will always be different, keeping the nodes heterogeneous. Table I below lists the simulation parameters considered.

TABLE I: NS2 SIMULATION PARAMETERS

| Parameter             | Value     |
|-----------------------|-----------|
| Network Size (x, y)   | 1300*1000 |
| Transmission Power    | 2.0W      |
| Receiving Power       | 1.0W      |
| Ideal Power           | 1.0W      |
| No. of nodes          | 40        |
| Packet size           | 1000bits  |
| Base Station location | 1400*1100 |

Fig. 2 shows the placement of nodes in the network area. Clusters are formed according to x-distances of the nodes. Node 40 is the sink node. From each cluster two nodes will be selected as cluster head having maximum residual energy among all nodes in cluster. As shown in Fig. 2, node contains in square and hexagon is the cluster heads of respective cluster. Base station node is shown in green color circle contain in red square box.

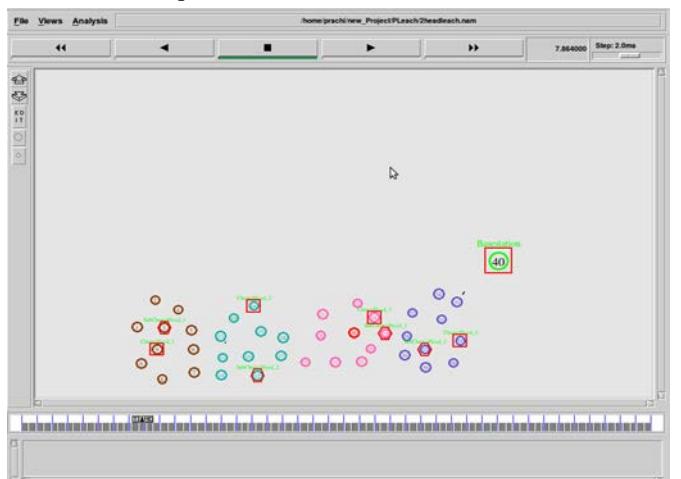


Fig. 2. Deployment of nodes.

Energy of each node is compared and node with highest energy is chosen as the cluster head. Second highest energy node is selected as deputy cluster head. This node will take

over the functionality of cluster head when present CH goes down. Fig .3 below shows selection of CH depending upon the energy level of the nodes in a cluster.

```

prachi@prachi-VirtualBox: ~/new_Project/PLeach
num_nodes is set 40
warning: Please use -channel as shown in tcl/ex/wireless-mitf.tcl
INITIALIZE THE LIST xListHead
Node 0 energy 100.000000
Node 1 energy 916.399840
Node 2 energy 932.106862
Node 3 energy 920.026492
Node 4 energy 885.248095
Node 5 energy 364.736184
Node 6 energy 121.039720
Node 7 energy 314.579225
Node 8 energy 133.039479
Node 9 energy 994.519187
Max energies 994.519187 Node numbers 9

```

Fig. 3. Cluster Head selection.

#### IV. RESULTS AND DISCUSSION

The network consisting of 40 heterogeneous nodes, deployed in an area of 1300 X 1000 and a base station located at 1400 X 1100, is considered for simulation purpose. 7 rounds are simulated here. Three protocols are considered here namely LEACH, PEGASIS and ours P-LEACH. The comparative results are discussed in this section. We considered energy consumption per round, mean delay per round and packet delivery ratio as the parameters for comparison.

##### A. Energy consumption

Energy consumption of the nodes depends on two tasks, i.e. communication and computation. Major energy is spent during communication, as its an energy intensive task. As compared to communication computation energy requirement is very very small. In this section, we will discuss about energy consumption in proposed protocol in comparison with LEACH and PEGASIS protocol.

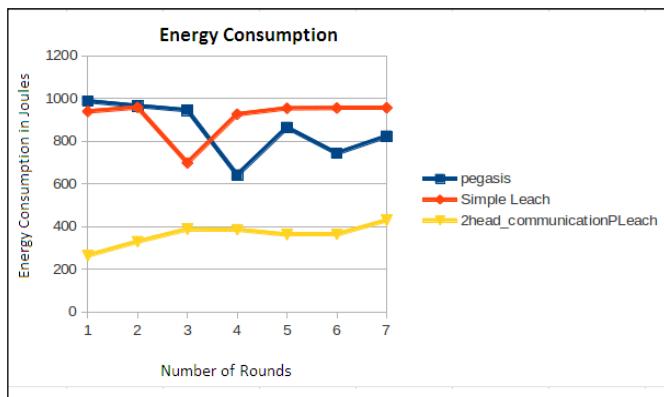


Fig. 4. Energy consumption in proposed method.

In the above Fig. 4 energy consumption in proposed P-LEACH protocol, LEACH and PEGASIS is calculated. As shown in figure, yellow line indicates energy consumption in proposed method, red and blue line indicates energy consumption in LEACH and PEGASIS protocols

respectively. From the graph it is clear that energy consumption in our proposed protocol is less as compare to LEACH and PEGASIS. This result is obtain by using same parameters for all three protocols. In case of LEACH direct communication to base station by each CH consumes more energy and cluster formation also consumes overhead energy. In PEGASIS leader node may be at a longer distance from BS, that consumes more energy. In case of proposed method, energy requirement for communication between nodes is less because CH selected will have maximum residual energy and distance between adjacent CH nodes is very less as compare to other protocols.

##### B. Packet delivery ratio (PDR)

Following Fig. 5 shows comparison of proposed method, LEACH protocol and PEGASIS protocol.

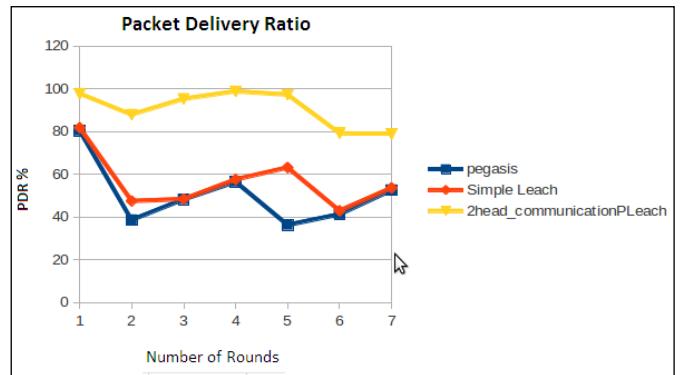


Fig. 5. Packet Delivery Ratio in communication.

In above graph,yellow line indicates the PDR in proposed method. Red and blue line indicates PDR in simpleleach and PEGASIS protocol respectively. Graph clearly indicates that in proposed method packet delivery ratio is more as compare to other two protocol. In this we have taken the interval 0.3-0.5sec to send packet from one node to another node. PDR in proposed protocol is coming out to be in the range of 80-100 which is good enough for network to become efficient network.

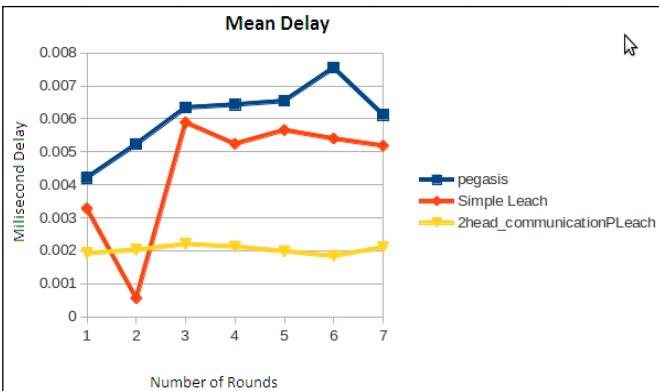


Fig. 6. Delay for packet sending.

Above Fig: 6 shows delay in the network. Yellow line indicates delay in proposed method where as red and blue lines indicate delay in LEACH and PEGASIS protocol repectivly. As shown in figure, in proposed method dealy is very less and it is in the range of 0.001msec to 0.002msec where as in case of LEACH and PEGASIS this delay vary in the range of 0.003msec to 0.0075msec. This result shows that delay in proposed method is very less in comparison with other two protcols which proves that network established using this protocol will be a very efficient network. Above discussed observations show that poposed method gives better result in terms of energy, packet delivery ratio and dealy. Also number of dead nodes in proposed method is less as compared to LEACH and PEGASIS protocol.

## V. CONCLUSIONS

The proposed method Improved P-LEACH protocol has been proven to produce a better result as compared to prior protocols such as LEACH and PEGASIS in WSN. Following are the key points which prove that proposed method is better as compare to LEACH and PEGASIS protocol.

*Minimum energy consumption:* In this protocol, energy required for communication between nodes is very less which results in improving network lifetime and network stability.

*Less PDR:* Packet delivery ratio is one of measure of judging network efficiency in WSN .In proposed method PDR is communing out to be 97.77% .Which means that there is very less packet drop while node communicates with each other. This results shows that there is less congestion in network which results in improving network efficiency.

*Less Delay:* In WSN, the network which has minimum delay that network will consider as more efficient network. Because, it takes less time to send packet from source to destination. In proposed protocol, delay is near about 0.009sec, which is less as compare to other two protocols. This shows that node takes minimum time to send packet from source to destination. This results in fast communication between nodes and minimum packet drop. Due to this network lifetime get increase.

*Number of Dead Nodes:* Dead nodes are responsible for breaking network communication and once the

communication is breaks that network will not be useful. In proposed method no. of dead is very less in whole network communication as compare to LEACH and PEGASIS. This result will increase the network's lifetime and efficiency of network.

## REFERENCES

- [1]. A.P. Chandrakasan, H.Balakrishnan and W.R Heinzelman, "Energy efficient communication protocol for wireless microsensor networks," *33rd Hawaii Int.Conf. Sys.Sci.*, Jan 2000.
- [2]. Lindsey, Stephanie, and Cauligi S. Raghavendra. "PEGASIS: Powerefficient gathering in sensor information systems." In *Aerospace conference proceedings, 2002. IEEE, vol.3*, pp. 3-1125.
- [3]. Al-Karaki, Jamal N., and Ahmed E. Kamal. "Routing techniques in wireless sensor networks: a survey." *Wireless communications, IEEE 11, no.6* (2004):6-28.
- [4]. Abdul Razaque, Musbah Abdulgader,Chaitrali Joshi, Fathi Amsaad, Mrunal Chauhan." P-Leach : Energy efficient routing Protocol" *IEEE,2016*
- [5]. Alekha Kumar Mishra, Rukhan Rahman,Rahul Bharadwaj,Rohit Sharma."An Enhancement of PEGASIS protocol with Improved Network Lifetime for WSN", *IEEE 2015*
- [6]. Agrawal, Palak, and P. R. Pardhi. "Routing Protocols For WSN." *International Journal of Computer Science And Applications 8, no. 1* (2015).
- [7]. Chenmin Li, Guoping Tan, Jingyu Wu, Zhen Zhang , Lizhong Xu "Analyzing Cluster-head Selection Mechanisms and Improving the LEACH " *2011 IEEE*.
- [8]. Naveen Kumar, Jasbir Kaur, "Improved LEACH Proto col for Wireless Sensor Networks." In *Proceedings of 7 th International Conference on Wireless Communications Networking and Mobile Computing*, Wuhan, China, 2011.
- [9]. Rashmi Jain, Manali Kshirsagar "Energy Saving in Heterogeneous Wireless Sensor Networks". *International Journal of Scientific & Engineering Research, Volume 4, Issue 8, August-2013*.
- [10]. Santar Pal Singh, S.C.Sharma " A Survey on Cluster Based Routing Protocol In Wireless Sensor Network" . *International Conference on Advanced Computing Technologies and Applications (ICACTA-2015)*.
- [11]. Padmavati, T.C. Aseri, "Performance Analysis of Cluster based Routing Protocols in Heterogeneous Wireless Sensor Network", *International Journal of Computer Applications (0975 – 8887) Volume 83 – No.16*, December 2013.
- [12]. Zhezhuang Xu, Chengnian Long, Cailian Chen, Xinpeng Guan "Hybrid Clustering and Routing Strategy with Low Overhead for Wireless Sensor Networks". *IEEE ICC 2010*.
- [13]. Rashmi Jain, Manali Kshirsagar, Latesh Malik "Analysis of Setup Energy of LEACH Protocol for Wireless Sensor Networks". *International Journal of Scientific and Engineering Research, June, 2016, Volume 7, Issue 5, May-2016 ISSN- 2229-5518*.
- [14]. Rashmi Jain, Manali Kshirsagar, Latesh Malik, "Heteroleach Protocol for Improvement of Stable Operation of Wireless Sensor Networks". *IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661, p-ISSN: 2278-8727, Vol 18,Issue 2*.

# On TCP-Induced Telehaptic Packet Loss and Jitter

Vineet Gokhale, Jan Fesl

Institute of Applied Informatics, University of South Bohemia, CZECH REPUBLIC, emails: {vgokhale, jfesl}@prf.jcu.cz

**Abstract:** Telehaptic data communication (transmission of touch signals) is known to be extremely sensitive to packet loss and jitter, the primary consequences of network congestion. Existing studies have established the Quality of Service (QoS) conditions that need to be guaranteed for smooth telehaptic communication. Specifically, the telehaptic communication can tolerate no more than 10% packet loss and 10 ms jitter. In this paper, we conduct a detailed investigation of the impact of TCP cross-traffic (pre-dominant traffic on shared networks) on telehaptic packet loss and jitter. The important contribution of our study is twofold. Firstly, we discover that even during scenarios where the long term average packet loss is comfortably below its QoS limit, the instantaneous loss can far exceed this limit. Secondly, we demonstrate that the probability of jitter QoS violation increases with the number of concurrent TCP sources in the network. These effects could potentially be harmful to the telehaptic activity, thereby raising serious concerns on designing efficient communication frameworks for minimizing telehaptic packet loss and jitter on shared networks.

**Keywords:** Telehaptic communication, QoS, shared network, packet loss, jitter

## I. INTRODUCTION

Everyday activities that the humans perform are largely dependent on our sensory mechanisms that aid in learning the physical properties of any real object such as size, shape, weight, texture, hardness, smell, and so on. The touch perception forms an integral part of our sensory mechanism. When an object is held, it exerts certain forces on the hand. The muscles and the joints of the hand capture these forces and they are then transmitted to the brain, generating a perception map of the object. This sensory mechanism is the fundamental driving force behind the innumerable forms of seamless interaction between humans and the physical world. Life would be lot harder if one was to light a matchstick, drive a car, or play a game of golf without the ability to feel the physical object.

*Haptics* relates to the science behind the different mechanisms of perception of real objects through the sense of touch. The deep research insights in this field have led to the design of elegant electro-mechanical systems that have enabled us to interact and manipulate virtual as well as remote objects through the feeling of touch.

*Telehaptic communication* – the science of coding, and subsequent transmission of haptic signals over a network – has witnessed rapid progress over the past decade. Such communication has been envisaged to redefine the way we interact with a remote world. For example, a surgeon could perform telesurgery on a distant patient through a robotic

telemanipulator with the feeling of touching the patient's body [1]. Telehaptic communication finds potential applications in a wide variety of other domains as well, like telemaintenance, and remote disaster management to name a few.

Figure 1 depicts a typical telehaptic communication framework over a shared network. The human operator (OP), using the force, audio, and video feedback from the remote environment, makes certain movements in an attempt to interact with and/or manipulate a remote physical object. The position and velocity signals thus generated are transmitted to the remote environment via the *forward channel*. The robotic teleoperator (TOP) at the remote location utilizes these coordinates in order to replicate OP's movements accurately. Any contact between the remote object and the TOP generates forces, which are transmitted back to the OP along with audio and video feedback on the backward channel. The presence of haptic feedback has been shown to increase the immersion into the remote environment, and further improve the precision of the telehaptic activity significantly [2].

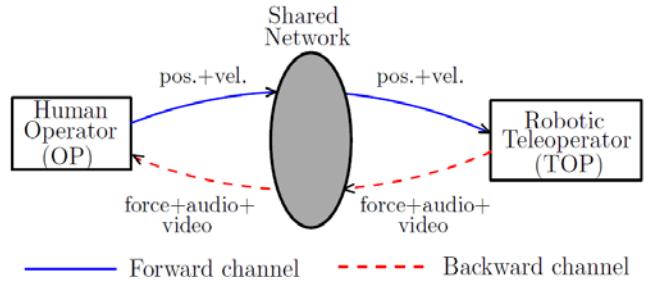


Fig. 1. Schematic representation of a telehaptic communication framework depicting the signal exchange between the human operator (OP) and the robotic teleoperator (TOP).

Naturally, such highly sensitive operations necessitate accurate replication of the OP's movements by the TOP, and also timely delivery of the feedback signals to the OP. For example, large delays in haptic feedback result in sluggish perception of the patient's body, thereby (potentially) leading to a wrong action by the surgeon. Additionally, large telehaptic jitter leads to perceiving the same remote object as having variable mass, which is absurd. Note that jitter refers to the variation in the packet delays. High packet losses may cause improper replication of the OP's movements accurately and/or OP being severely deprived of the feedback signals. Both these scenarios could have catastrophic effects on the ongoing telehaptic activity. Note that the packet loss in the network is a consequence of queue overflows during congestion. These effects can, at times, cause irreparable damage to the patient. Hence, the communication network that transfers the telehaptic feedback plays an instrumental role in determining the quality of the telehaptic interaction.

Experimental studies, such as [3], have demonstrated that the human perception of haptic feedback can tolerate a

maximum packet loss and jitter of not more than 10% and 10 ms, respectively. This means that the perception of the remote environment is not hampered even if at least 90% of the telehaptic samples reach the OP/TOP with a jitter of no more than 10 ms. These telehaptic packet loss and jitter constraints that need to be satisfied for a seamless telehaptic activity are collectively known as *Quality of Service (QoS)*. For a smooth telehaptic activity, the network needs to guarantee QoS-compliance at all times. In general, QoS violations lead to deteriorated perception of the remote environment, as explained previously.

It is important to note that the work in [3] treats the packet loss as a time-average entity. In other words, the work in [3] averages packet losses over an entire telehaptic session; the authors discovered that when this long term average packet loss exceeds 10%, the users started perceiving an unacceptable deterioration in the perception of the remote environment. Note that the long term average packet loss refers to the average of the packet loss measured over the entire duration of the telehaptic session. It is worth noting that this work does not consider the characteristics of the instantaneous loss while establishing 10% as the packet loss criteria for smooth perception.

It is important to remark that in a real world scenario the perception of remote objects (potentially) depends on the instantaneous packet loss rather than the long term average loss. For example, in a few network settings the instantaneous packet loss is way higher than 10% (see Figures 4 and 5) despite its long term average value being below 10%. This means that a vast majority of the telehaptic samples (up to 80% in our simulations) do not reach the destination. As per the claim in [3], this implies that even when all packets (100%) are lost over a certain interval, the users do not feel any perceptual degradation. This is incorrect as no haptic feedback leads to improper perception of the remote world. Therefore, the instantaneous packet loss, and not the long term average loss, is a more relevant performance metric from the standpoint of perception in any telehaptic communication.

A telehaptic stream on a shared network, like the Internet, has to contend with other cross-traffic streams that are concurrently being served by the network. Hence, it is crucial to study the influence of the coexisting cross-traffic streams on the telehaptic stream in terms of instantaneous packet loss and jitter. On a shared network, the telehaptic stream is guaranteed to encounter Transmission Control Protocol (TCP) traffic since TCP amounts to over 90% of the overall traffic [4]. TCP provides a reliable data communication mode, and hence forms the cornerstone of a wide variety of Internet services that require reliable transfer of data, such as email, file transfer, web browsing, and video streaming applications like YouTube, and Netflix.

For our investigation in this paper, we consider a specific flavor of TCP named TCP NewReno [5]. TCP NewReno (or any TCP source in general) is a rate-adaptive transport layer protocol that controls its transmission rate depending on the congestion level in the network. The TCP source uses packet loss as an indicator of congestion. Based on the packet loss as detected by the source, the data rate is adapted to match the available network bandwidth, and thereby eschew

underutilization the network resources. The TCP source increases its data rate until it detects a packet loss (indicating congestion). In response, it reduces the data rate in order to relieve the network, and thereby achieve congestion control. Once the source detects that the network is free, it begins to increase the data rate, and this cycle continues. As can be observed, the TCP source relies heavily on the packet loss in the network in order to learn the available network bandwidth. In fact, the working principle of TCP is itself based on inducing packet loss in the network. This behavior naturally impacts the concurrent streams in the network. In addition, the data rate variation of TCP also introduces jitter that negatively affects the telehaptic activity. In this work, we are interested in studying whether these packet loss and jitter effects of TCP have any notable impact on QoS-compliance of the telehaptic stream.

In this paper, we intend to study the impact of multiple TCP cross-traffic sources on a telehaptic stream. The objective of this investigation is to gain insights into the characteristics of the instantaneous telehaptic packet loss and jitter under the influence of coexisting TCP cross-traffic sources. The contribution of our work is as follows.

- (i) We demonstrate that in a wide range of settings, even though the long term average packet loss meets the QoS criteria, the instantaneous packet loss can be much higher.
- (ii) We show that the peak telehaptic jitter can far exceed the 10 ms deadline for standard network settings, and hence is extremely prone to QoS violations.

The remainder of the paper is organized as follows. In Section II, we discuss in brief a few prior works available in the literature related to the interplay between TCP and telehaptic streams. Section III describes the detailed simulation setup that we designed for our investigation. In Section IV, we present the results of our experiments, and in Section V, we state our conclusions and mention potential directions for future research.

## II. RELATED WORK

Only a handful of works have attempted to study the behavior of telehaptic streams on a shared network [6, 7, 8]. Although these works considered network cross-traffic in their performance evaluation, negligible attention is paid to the TCP streams that form a major component of the overall cross-traffic. A recent work [9] conducted a comprehensive analysis of the effects of a single TCP stream on the long term average telehaptic packet loss as well as jitter. However, this work investigates ignores the instantaneous packet loss. As explained earlier, the instantaneous packet loss forms a more important performance metric than the long term average measure. Furthermore, this analysis confines the number of concurrent TCP streams to one. Hence, the effect of multiple TCP streams on telehaptic loss and jitter remains unexplored in this work.

## III. SIMULATION SETUP

In this section, we give a detailed description of the experimental settings considered in our simulations. The goal of this section is to develop an understanding of the dynamics of interplay between TCP and telehaptic streams when the two

traffic types share a single bottleneck link. We carry out our investigation using NS3 – a discrete event network simulator [10]. We use the single bottleneck network topology as shown in Figure 2.  $H_1$  and  $H_2$  are the OP and the TOP, respectively, of the telehaptic communication framework shown in Figure 1.  $[S_1, \dots, S_n]$  and  $[R_1, \dots, R_n]$  are the sets of  $n$  TCP sources and receivers, respectively.  $L_1$  is the bottleneck link on the forward channel. Note that the data rate variation of TCP influenced the queue occupancy at  $B_1$ , the router at the ingress of  $L_1$ . As mentioned earlier, the TCP sources employ NewReno congestion control scheme. For telehaptic communication, we leverage the protocol proposed in [8]. It can be shown that in presence of TCP NewReno sources, a telehaptic source employing the protocol in [8] generates packets at the rate of 250 per second. The packet scheduling at the network queues is based on the standard drop-tail mechanism.

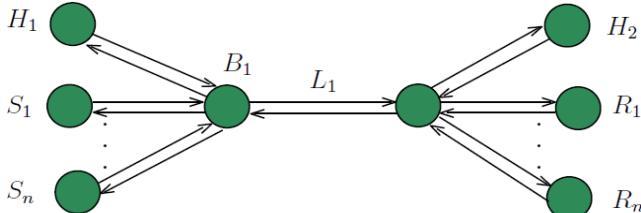


Fig. 2. Single bottleneck network topology used in our simulations.

Notations:  $H_1$  and  $H_2$  – operator and teleoperator in telehaptic communication, respectively;  $[S_1, \dots, S_n]$  – TCP sources;  $[R_1, \dots, R_n]$  – TCP receivers;  $L_1$  – bottleneck link;  $B_1$  – router at the ingress of bottleneck link.

The propagation delay of each link is set to 5 ms, and hence the one-way propagation delay between a source and its corresponding receiver is 15 ms. The channel capacity of  $L_1$  is set to 3 Mbps. The access links to  $L_1$  have high capacities of 5 Gbps. The queue size at the  $B_1$  is configured to 15 kB. The TCP and the telehaptic packets have sizes 578 B and 512 B, respectively, unless mentioned otherwise. For the purpose of our simulations, we consider  $n$  in the range [1, 10]. However, it is worth remarking that the observations that we make regarding the telehaptic loss and jitter hold good for higher values of  $n$  as well.

All sources begin the transmissions simultaneously at  $t = 0$ . We run each simulation until  $t = 100$  s. Throughout the simulations, we record the packet loss and jitter encountered by the telehaptic sources.

#### IV. RESULTS

In this section, we present the results of our investigation of telehaptic packet loss and jitter induced by the coexisting TCP streams. We begin by reporting the packet loss, and then move to the jitter part.

In Figure 3, we report the long term average packet loss seen by the telehaptic source by varying  $n$  over the considered range. It can be seen that the long term average packet loss is an increasing function of  $n$ . However, for  $n < 10$ , the long term average packet loss complies to the QoS limit of 10%. However, we note that for higher  $n$ , the average loss exceeds the QoS limit severely. For brevity, we do not report the telehaptic packet loss in the higher  $n$  regime.

We now move to studying the behavior of the instantaneous telehaptic packet loss for a specific value of  $n$  for which the average loss meets the telehaptic packet loss criteria. For this purpose, we choose  $n = 10$ . Note, from Figure 3, that the long term average loss for  $n = 10$  is approximately 10%. From Figure 4, it can be seen that the instantaneous telehaptic packet loss varies rapidly between 0 and 50%. In addition to the peak loss measurement of 50%, it can also be seen that the packet loss QoS criteria gets violated regularly. Although we report the instantaneous packet loss only for  $n = 10$ , we observe similar behavior for other values of  $n$  as well. In short, even though the long term average packet loss meets the QoS criteria, the instantaneous loss can be significantly higher. This confirms our conjecture that the instantaneous packet loss should be considered as the performance metric rather than the long term average packet loss.

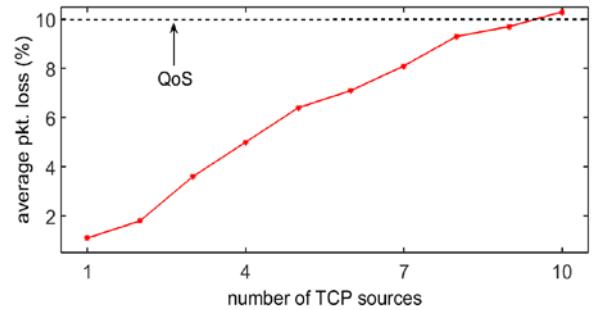


Fig. 3. Evolution of long term average telehaptic packet loss as a function of the number of TCP sources.

It is important to remark that even though the interval over which the QoS violation occurs is small (a maximum of 300 ms), this could potentially have severe artifacts considering the scale of sensitivity that a telehaptic activity, like telesurgery, requires.

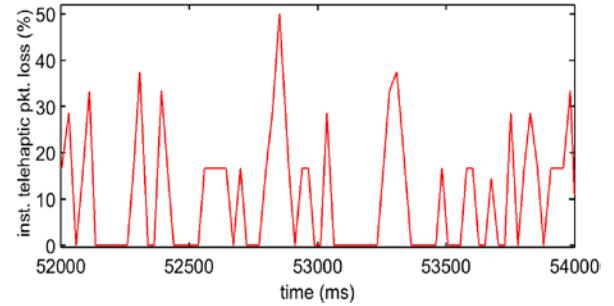


Fig. 4. Instantaneous telehaptic packet loss  $n = 10$  showing significant overshoot compared to its long term average value.

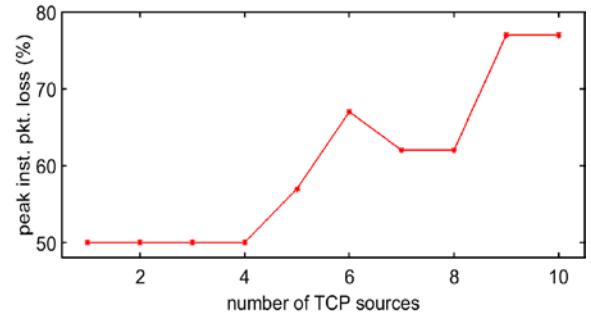


Fig. 5. Variation of peak instantaneous telehaptic packet loss as a function of the number of TCP sources in the network.

Having seen the instantaneous loss, we now turn towards determining the peak instantaneous telehaptic packet loss in the simulations. Figure 5 shows the variation of the peak packet loss in the considered range of  $n$ . It can indeed be observed that the instantaneous packet losses are substantially higher despite the long term average packet loss complying to the QoS requirement. Therefore, we demonstrate through experiments that any guarantees on the long term average packet loss do not imply any guarantees on the peak instantaneous packet loss. This suggests that in order to ensure a seamless telehaptic activity, one must design communication frameworks that can provide QoS guarantees on the instantaneous telehaptic packet loss.

It has been shown in the past that smaller telehaptic packets (relative to TCP packets) are less susceptible to losses [9]. Specifically, the experiments in [9] reveal that the telehaptic packets of size 137 B are rarely dropped by the network queues in presence of a single TCP source that transmits packets of size 578 B. Hence, one potential solution for mitigating the telehaptic losses is to minimize the packet sizes. The other plausible remedy could be to design priority queueing schemes that can serve packets carrying crucial telehaptic data with higher precedence over other cross-traffic streams.

We now move to the peak telehaptic jitter measurements. For the packet sizes mentioned in Section III, we notice that the peak telehaptic jitter varies in the range [2.55, 6.69] ms, which satisfies the QoS constraint. Since the jitter is known to be heavily dependent on the TCP packet size, for concreteness in exposition, we run the simulations with TCP packets of size 1042 B, which is also another standard value. In Figure 6, we plot the peak instantaneous telehaptic jitter as a function of  $n$ . It can be seen that the peak jitter is a non-decreasing function of  $n$ . Further, for  $n > 5$ , the jitter QoS condition is severely violated. This implies that higher the number of concurrent TCP streams, larger is the probability of violation of the telehaptic jitter QoS violation.

Based on the analysis in [9], it is reasonable to argue that the peak instantaneous haptic jitter increases with the number of concurrent TCP streams as well as their packet sizes. For QoS-compliance, one needs to theoretically determine upper bounds for these two factors. The network administrator then needs to ensure that the cross-traffic satisfies the two bounds. This guarantees satisfaction of QoS conditions for telehaptic jitter.

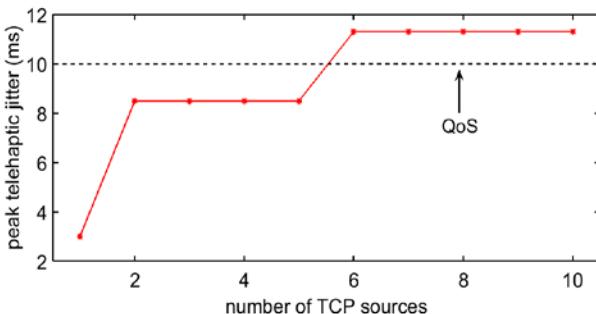


Fig. 6. Variation of peak instantaneous telehaptic jitter with TCP packets of size 1042 B.

## V. CONCLUSIONS

In this paper, we conducted an extensive investigation of the interplay between TCP NewReno and telehaptic streams. We demonstrated that even though the long term average telehaptic packet loss satisfies the QoS criteria, the instantaneous loss can far exceed the QoS limit of 10%. Additionally, we showed that the telehaptic stream faces extreme jitter QoS violations for TCP packets of standard sizes. Hence, we conclude that it is crucial to monitor and control the number of TCP streams, as well as the size of TCP packets in order to achieve seamless telehaptic communication on a shared network.

In a future version of this article, we intend to propose a telehaptic communication framework that mitigates the detrimental effects of TCP sources. Also, studying the effects of other variants of TCP on telehaptic stream could be another interesting avenue for future research.

## REFERENCES

- [1] R. Anderson, and M. Spong, "Bilateral control of teleoperators with time delay," *IEEE Transactions on Automatic Control*, vol. 34, pp. 494–501, May 1989.
- [2] C. Basdogan, C. Ho, M. A. Srinivasan, and M. Slater, "An experimental study on the role of touch in shared virtual environments," in *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 7, pp. 443–460, 2000.
- [3] P. Dev, D. Harris, and D. Gutierrez, A. Shah, and S. Senger, "End-to-end performance measurement of internet based medical applications," in *Proceedings of the Annual Symposium of the American Medical Informatics Association*, Nov 2002.
- [4] S. Ryu, C. Rump, and C. Qiao, "Advances in internet congestion control," in *IEEE Communications Surveys and Tutorials*, vol. 5, pp. 28–39, 2003.
- [5] S. Floyd, A. Gurto, and T. Henderson, "The NewReno modification to TCP's fast recovery algorithm," 2004.
- [6] R. Wirz, M. Ferre, R. Marin, J. Barrio, J. M. Claver, and J. Ortego, "Efficient transport protocol for networked haptics applications," in *Haptics: Perception, Devices and Scenarios*, Springer, 3–12, 2008.
- [7] V. Gokhale, J. Nair, and S. Chaudhuri, "Opportunistic adaptive haptic sampling on forward channel in telehaptic communication," in *Haptics Symposium*, Apr 2016.
- [8] V. Gokhale, J. Nair, and S. Chaudhuri, "Congestion control for network-aware telehaptic communication," in *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 13, 2017.
- [9] V. Gokhale, J. Nair, and S. Chaudhuri, "Teleoperation over a shared network: When does it work?," in *International Symposium on Haptic, Audio and Visual Environments and Games*, Oct 2017.
- [10] "NS3 - The network simulator", 2011, online - <http://www.nsnam.org/>.

# Specialized Computer Systems

# Machine Learning Methods in Electrocardiography Classification

Iryna Mykoliuk<sup>1</sup>, Daniel Jancarczyk<sup>1</sup>, Mikolaj Karpinski<sup>1</sup>, Viktor Kifer<sup>2</sup>

1. Department of Computer Science and Automatics, University of Bielsko-Biala, POLAND, Bielsko-Biala, 2 Willowa St,  
email: i.mykoliuk@gmail.com, djancarczyk@ath.bielsko.pl, mkarpinski@ath.bielsko.pl

2. Faculty of Computer Information Systems and Software Engineering, Ternopil Ivan Puluj National Technical University, UKRAINE,  
Ternopil, 56 Ruska St, email: kifervictor@gmail.com

**Abstract.** This document reports results that tend to confirm the applicability of the machine learning combined with signal processing for automatic atrial fibrillation detection from a short single lead electrocardiography recording.

**Keywords:** machine learning, signal, ECG, classification, heartbeat.

## I. INTRODUCTION

Automated electrocardiography (ECG) analysis has a number of ground tasks which include noise removal, QRS detection, P and T waves detection etc. The first two problems already have a number of techniques which provide good results. In particular band pass filters, Fourier based analysis and transforms and wavelet transformations are commonly used for noise reduction. Pan Tompkins algorithm, various transforms like Wavelet, Hilbert and Empirical Mode Decomposition combined with some decision logic can identify QRS complexes with detection rates over 99%.

P and T wave have small amplitudes, which makes them less noise resistant. But position and shape of the waves are the important components of ECG analysis. With the most recent approaches based on advanced Kalman filters and wavelet transforms, detection accuracy for those waves reached over 90%.

In general, ECG classification solutions tend to be class testing, when a particular record is checked against some abnormal rhythm pattern. This happens because there's a large variety of ECG shape changes, which are hard to handle by a single algorithm. And there is a strong tendency in last decade of moving from a threshold-based analysis (which suffers from introducing to validation dataset ECG recordings from new patients or recorded with new devices) to data-driven approaches which include classical supervised machine-learning models and neural networks.

## II. BIOLOGICAL BACKGROUND

Understanding of heart activity is an important part. Human heart consists of four chambers: left atrium (LA), right atrium (RA), left ventricle (LV) and right ventricle (RV). Between RA and RV there is a tricuspid valve, and between LA and LV there is a mitral valve. These valves prevent blood going directly from atriums to ventricles. Ventricles also have valves which prevent blood from going into veins, there is a pulmonary valve in right ventricle and an aortic valve in the left one.

### Electrocardiogram

ECG is the process of recording the electrical activity of the heart over a period of time using electrodes placed on the skin. These electrodes detect the tiny electrical changes on the skin that arises from the heart muscles.

In a conventional 12-lead ECG, 6 electrodes are placed on the surface of the chest, 2 on hands, and 2 on legs. The overall magnitude of the heart's electrical potential is then measured from 12 different angles ("leads") and is recorded over a period of time. A healthy heart has a specific order of polarization and depolarization during each heartbeat. It starts with the sinoatrial node, then spreads through the atrium to atrioventricular node, and then to ventricles.

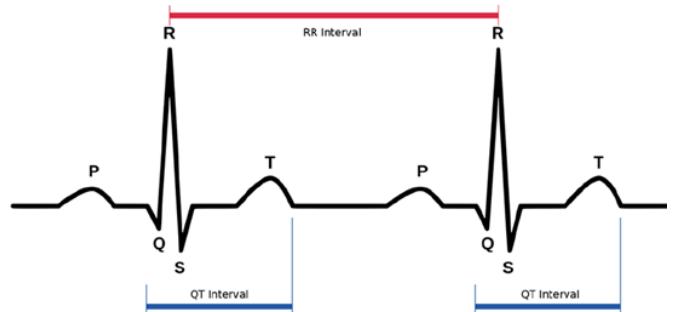


Fig. 1.Typical ECG trace

The graph of voltage versus time produced by this noninvasive procedure is called electrocardiogram. Each ECG consists of periodic PQRST complexes which represent a cycle of heart activity. In particular, one complex has:

1. P-wave (atrial contraction)
2. QRS-complex (contraction of the ventricles)
3. T-wave (relaxation of the ventricles)

The duration of PQRST might differ depending on the heart rate. There are a couple of ways how to measure it, one of the most common is called RR-interval, which represents the duration as a distance between two following R peaks

### Common abnormal heart rhythms

Heart irregular rhythm can be classified by heart rate into tachycardia (heart beats too fast, more than 90 bpm) and bradycardia (heart beats too slow, less than 60 bpm); and by place of occurrence into supraventricular (atria contracts irregularly) and ventricular (ventricles contract in an irregular pattern) arrhythmias.

Atrial fibrillation is an abnormal heart rhythm characterized by rapid and irregular beating and which has no symptoms in most of the cases. It is the most common serious abnormal heart rhythm and it affects approximately 2-3% of the population in Europe and North America. Also, the percentage of people with AF increases with age with 0.14% under 50 years old, 4% between 60 and 70 years old and 14% over 80 years old being affected. On ECG, atrial fibrillation is usually diagnosed by absence of P-wave and irregular heart beats pattern.

Atrial flutter (AFL) is an abnormal heart rhythm, similar to atrial fibrillation. Both of them are types of supraventricular tachycardia. In AFL the electrical signal goes along the pathway in a circular motion, which results in atrial muscle contractions that are faster than and out of sync with the lower chambers (ventricles). Similar to A-Fib the heart beats fast, but in a regular pattern.

Itself atrial flutter is not life-threatening, but due to slower blood pumping it creates, the side effects might cause health problems.

On ECG atrial flutter might be diagnosed by the presence of multiple f-waves instead of a P-wave.

Paroxysmal supraventricular tachycardia (PSVT) is characterized by the episodes of a rapid increase of heart rate (100-250 bpm) and usually regular heart rhythm. It is most often seen in young people and infants. Alcohol, illicit drugs, caffeine, and smoking might be a cause of this arrhythmia.

Another abnormal heart rhythm is premature ventricular complex (PVC) characterized by long QRS complexes (> 0.12ms), ST segment and T wave changes (irregular, inverted in leads with a dominant R wave, etc).

Ventricular fibrillation (V-Fib) is a life-threatening abnormal heart rhythm when the lower chambers quiver and the heart can't pump blood to the body. This occurs because of electrical signal disorder which makes chambers contract very fast in an unsynchronized way. It requires immediate treatment and can usually be stopped with a defibrillator.

### III. DATA PREPROCESSING

#### *Handling of imbalanced data*

The data in selected dataset is quite imbalanced which might create problems during training. The model might easily overfit and predict all of the time the most represented category. This creates a so-called accuracy paradox: suppose that we have a dataset with two labels A and B and 90% of the records are instances of A, the classifier might predict all the time class A which will give the accuracy of 90% although it definitely won't be a good classifier. There are several common approaches here: use abnormalities detection algorithms or apply class weights or perform data balancing.

The first approach was applying class weights as a logarithmic function of the proportion of total instances divided by the number of class instances.

$$\text{classweight} = \min\left(\log\left(\frac{\mu * \text{totalinstances}}{\text{classinstances}}\right), 1\right)$$

where  $\mu$  - parameter to tune

While applying class weight helped to prevent overfitting on early training stages, the model still failed to learn class-specific features.

The second approach was to apply data balancing. There are many techniques available for performing data balancing including:

Under-sampling (deleting instances of over-represented classes)

Over-sampling (repeating training with under-represented classes)

Generating Synthetic Samples (generating new samples of the under-represented based on available samples)

Use algorithms which are better in handling imbalanced data.

Under-sampling, over-sampling and using of decision trees which can handle imbalanced data were used to work out the problem.

#### *Normalizing data*

Measuring ECG values might generate a wide range of values depending on different conditions when the recording was performed. This might misdirect model to learn absolute values of one instance instead of the value for all instances.

To prevent this situation, data normalization is applied which means adjusting measured values to a common scale. The first step is to find a baseline (mean value of the record) and subtract it from the record values. And after that divide record values by the absolute value of the record. So the final scale would be in the range [-1; 1].

#### *Band pass filter*

ECG signal might be altered with the noise of different sources. This might be electrical signals coming from human body activity (usually electromyography signals from muscles contraction), procedure artifacts (eg. connecting/disconnecting electrodes), power line interference and instrumentation noise.

Power line interference is not less represented in hand-held devices, and usually, occurs through inductive mechanism. The power lines across the world have the frequency of 50Hz or 60Hz, which is higher than the subject information frequencies (0.1Hz to 40Hz). So it can easily be eliminated by filters as presented on Fig. 2.

Both electrode contact noise and electrode motion noise cause baseline changes that occur due to variation in the position of the heart with the respect to the electrodes. More specifically, the amplitude of the changes is defined by variation of electrode-skin impedance (ESIV). The larger ESIV, the smaller change of impedance needed to cause a shift in the baseline. This kind of noise is not defined by any particular range of frequencies and it isn't easy to eliminate it with signal processing, the task is to perform the recording procedure without patient disturbance.

Electromyography (EMG) noise is caused by contraction of muscles, not related to the heart activity (moving of hands and/or legs). For the hand-held devices usually, the highest EMG noise component is waves generated by depolarization and repolarization of hand muscles (as ECG signal goes from heart to the device directly through hands). EMG signal frequency is stochastic in nature, but the significant activity

happens in the range of 5Hz and 450Hz. This partially overlaps with the frequencies of the ECG signal so it's hard to eliminate it absolutely. In particular, the problematic areas are P and T waves of the ECG.

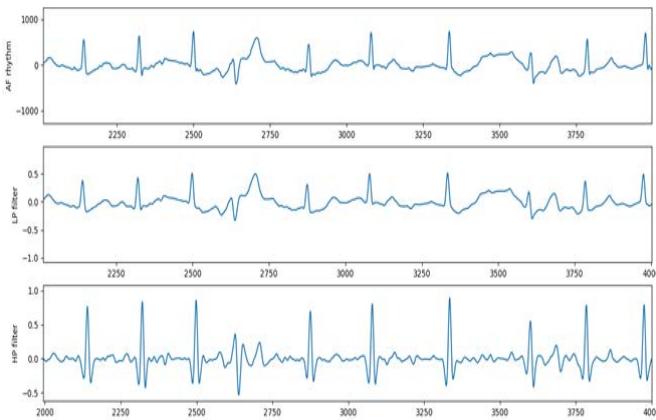


Fig.2. Signal preprocessing with low-pass and high-pass band filters

So to remove most of the noise the combination of low pass and high pass bands filters were used. The range of frequencies to be kept is from 0.1Hz to 40Hz.

The SciPy filter function in combination with NumPy convolve function was used to implement both filters.

#### IV. FEATURE SELECTION

Feature selection includes:

1. QRS detection.
2. Heartbeat extraction.
3. Features extraction

The task of QRS detection is to find R-peaks which represent the heart beats. Beat detection is a procedure preceding any kind of ECG analysis and so it is critical to correctly detect heart beats.

One of the standard algorithms used for QRS detection is a Pan-Tompkins algorithm. The original paper states detection rate of 99.3%. Many modifications were made to the algorithm since its publication including additional filtering, search back support, outliers rejection etc.

The filtered signal is introduced as an input to the algorithm. The processing of the data starts with a first-derivative filter, which helps obtain information about the slope of the QRS.

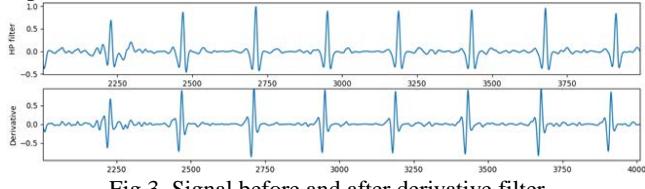


Fig.3. Signal before and after derivative filter

To get even better intensity of the slope, the squaring with normalization is applied. This gives a graph where most of the values are near to zero, and in the positions of R peaks, it has a value much higher (usually over 0.5).

Additionally moving average is applied, which combines two close peaks on the graph above into one. After that, the algorithm finds positions on the graph that are higher than the average (referenced threshold). And for all of those positions

it looks for the position of the higher value. This position is marked as potential R-peak location.

R-peaks then go through false rejection algorithm which checks the average length of the RR-interval, and rejects those R-peaks, where the length of RR-interval is less than half of the average.

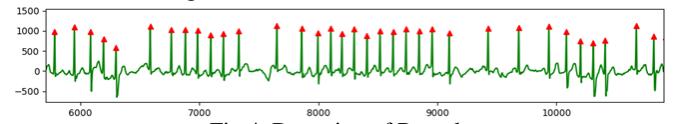


Fig.4. Detection of R-peaks

There is also additional search back algorithm which again looks on RR-intervals and finds positions where there might be a missed R-peak (RR-interval is higher than 1.5 of the average). There it lowers the threshold and runs the search algorithm once again.

Heartbeat extraction is done with the static window. By taking in consideration that normal heart rhythm is in the range of 60 to 100 beats per minute, we considered static windows of sizes from 0.5s (120 bpm) to 0.7s (86 bpm). The window size that served best was 0.6s with the distribution of 0.15s for P-wave, 0.1s for QRS complex and 0.35s for ST interval and T-wave.

For each R-peak detected, the heartbeat template was extracted with the static window of 0.6s which includes 0.2s before and 0.4s after R-peak.

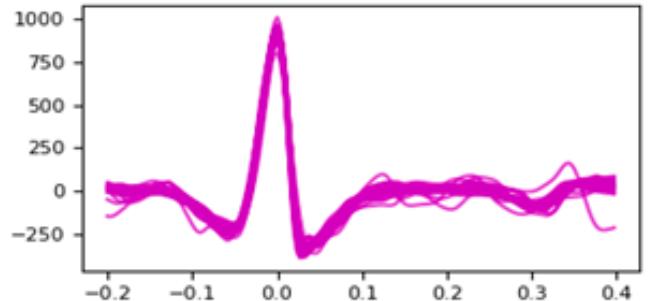


Fig. 3. Extracted heartbeat

During the model development process, many features were considered as good for prediction. To analyze the noise of the data, transformation coefficients were used. In particular, first 400 coefficients of Fourier transformation and details coefficients of 4-level wavelet decomposition.

Another range of useful features is heart rate variability. It is computed on RR-intervals, which includes time domain components (RMSSD, SDNN, NN20, PNN20, NN50, PNN50, mean RR, std RR, mean HR) and frequency domain components.

Frequency domain components are computed for 3 frequency bands: very low frequencies (0-0.04), low frequencies (0.04-0.15) and high frequencies (0.15-0.4). To compute frequencies powers, Welch estimation of power spectral density is performed on interpolation of RR-intervals, and the total values of power bands are computed by integrating (with composite trapezoidal rule) function of frequencies and powers.

Additionally, minimum and maximum RR-interval lengths are added to the feature vector.

The extractor also computes the proportion of the number of R-peaks detected divided by the length of the record, mean and standard deviation of normalized R-peak values, number, and proportion of detected R-peak values with the inverse sign.

Based on the templates of heart beats the module finds a median heartbeat as it best represents the most common template of heartbeat. Median was used rather than mean because it is more noise resistant. The extractor also computes the standard deviation of each point of templates.

## V. CLASSIFICATION

One of approaches to ECG classification was based on decision trees. The bagging was used from the very beginning with the random forest algorithm.

Random forest implies divide-and-conquer principle, which makes it easy to execute in parallel. This makes training process very fast (a couple of seconds). This is particularly good for experiments with the trees structures optimizations. For classification process was selected Random forest algorithm which consists of 60 decision trees. The implementation of the algorithm is taken from the Scikit-learn library. Results of classification process for random forest based model shown in table 1.

TABLE 1. SCORES FOR RANDOM FOREST MODEL.

| Class identifier | Precision | Recall | f1-score |
|------------------|-----------|--------|----------|
| 0                | 0.77      | 0.67   | 0.72     |
| 1                | 0.73      | 0.75   | 0.74     |
| 2                | 0.64      | 0.68   | 0.66     |
| 3                | 0.68      | 0.61   | 0.65     |
| Avg/total        | 0.70      | 0.70   | 0.70     |

Another approach was based on fully convolutional network, this type of network architecture serves as a feature extractor, and the final decision is made by a fully-connected layer. Results of classification process for neural network based model shown in table 2.

TABLE 2. SCORES FOR NEURAL NETWORK MODEL.

| Class identifier | Precision | Recall | f1-score |
|------------------|-----------|--------|----------|
| 0                | 0.58      | 0.26   | 0.36     |
| 1                | 0.65      | 0.94   | 0.77     |
| 2                | 0.60      | 0.28   | 0.38     |
| 3                | 0.47      | 0.03   | 0.06     |
| Avg/total        | 0.62      | 0.64   | 0.59     |

Model used 4 categories ‘0’ – is atrial fibrillation, ‘1’ – is

normal sinus rhythm, ‘2’ – is other heart rhythm and ‘3’ – is noisy signal. Precision is defined as the number of true positives divided by the number of true positives and false positives and is a measure of a class exactness. Recall is the number of true positives divided by the sum of a number of true positives and false negatives and it expresses the measure of classifier completeness. F1 score is used as a model estimate. F1 measure is defined a combination of specificity and sensitivity:

$$F1 = \frac{2 * precision * recall}{Precision + recall}$$

It can be used to express the average of classifier exactness and completeness. Scoring algorithm include F1 metrics only for 3 categories: atrial fibrillation, normal rhythm, and other rhythm:

$$F1 = \frac{F1(AF) + F1(N) + F1(O)}{3}$$

So the final score of the model is an average of F1 scores for 3 classes.

## VI. CONCLUSIONS

This paper shows the ECG automated analysis problem and one of suitable algorithms for this – Random forest. Was described the main steps of ECG signal preprocessing which includes handling of imbalanced data, data normalization and noises filtering. Also presented QRS complexes detection and the main features for ECG classification.

## REFERENCES

- [1] Heart Disease Facts // <https://www.cdc.gov/> URL: <https://www.cdc.gov/heartdisease/facts.htm> (last accessed: 2018/03/28).
- [2] Wang, Zh., Yan, W., Oates, T. Time Series Classification from Scratch with Deep Neural Networks: A Strong Baseline // *arXiv preprint. arXiv:1611.06455*, 2016.
- [3] Brownlee J. Classification accuracy is not enough: More performance measures you can use // <http://machinelearningmastery.com/> URL: <http://machinelearningmastery.com/classification-accuracy-is-not-enough-more-performance-measures-you-can-use> (last accessed: 2018/03/28).
- [4] ECG basics // <https://www.osmosis.org/> URL: [https://www.osmosis.org/learn/ECG\\_basics](https://www.osmosis.org/learn/ECG_basics) (last accessed: 2018/03/28).
- [5] Johnson, A. E., Behar, J., Andreotti, F., Clifford, G. D., & Oster, J. (2014, September). R-peak estimation using multimodal lead switching. In *Computing in Cardiology Conference (CinC)*, 2014 (pp. 281-284). IEEE. Thomas Mitchell. Machine Learning / Thomas Mitchell. – McGraw-Hill, New York, 1997. – ISBN:0070428077.

# Construction of Vertical Scanner for Laser Analysis of Gel Samples

Jiří Pech<sup>1</sup>, Milan Novák<sup>2</sup>, Ladislav Ptáček<sup>3</sup>, Jana Kalová<sup>4</sup>

Faculty of Science, University of South Bohemia, Branišovská 1760, 370 05 České Budějovice, Czech Republic, e-mail: [pechj@prf.jcu.cz](mailto:pechj@prf.jcu.cz)<sup>1</sup>, [novis@prf.jcu.cz](mailto:novis@prf.jcu.cz)<sup>2</sup>, [lptacek@prf.jcu.cz](mailto:lptacek@prf.jcu.cz)<sup>3</sup>, [ikalova@prf.jcu.cz](mailto:ikalova@prf.jcu.cz)<sup>4</sup>

**Abstract:** This paper is focused on the construction of vertical scanner with linear translation used for gel samples analysis via laser. Scanner construction is based on the construction of RepRap 3D printer. The paper describes the process of choosing this type of construction and then the construction itself. Advantages and disadvantages of this solution are recorded too.

**Keywords:** Industry 4.0, 3D prints, Arduino, Raspberry Pi, G-Code.

## I. INTRODUCTION

This paper describes the design, the construction and the testing of the vertical scanner, which was built for the Institute of Physics and Biophysics [1]

The construction is based on the construction of the 3D printer Rep Rap Rebelix (figure 1) founded on the Rambo motherboard, supported by the Marlin firmware. All of the components are open software or open hardware, so we could use them and modify them as we needed.

The single-board computer Raspberry Pi with the attached touch display is used as the control computer. The control program is written in the Python language.

It means, that this article describes the applied research.

## II. TASK

Institute of Physics and Biophysics [1] assigned us following task: "Design a vertical scanner with linear translation that will be used for gel samples analysis via laser. The scanner has to consist of a frame, which will be moving in two directions: left - right and up - down. But in the direction backwards – forwards it will be stable. Next task is to cover the sample fully and to not stop at any position." The Physicists searched the Internet, but they haven't found any satisfying solution. Only one of the found solutions was little bit fair but it was very expensive (about 1000 Euro). In addition, this scanner's recommended working position is horizontal, and we need the vertical working position because our lasers have horizontal rays.

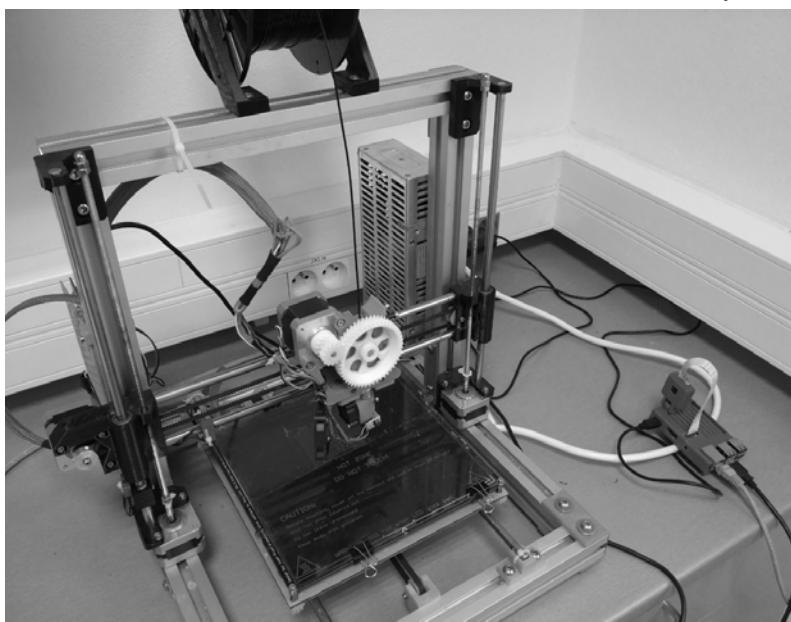


Fig. 1. 3D printer Rep Rap Rebelix with the control computer Raspberry Pi (on the right side).

So, it was clear, that if we want to satisfy them, we have to design absolutely new construction.

## III. METHODS

After the acceptance of this challenge, we started thinking about the construction of the scanner. First, we thought about using linear motor, but we discovered the fact that all of the

available linear motors are very expensive. So, we decided to use a motor with a screw-thread.

We have the 3D printer Rep Rap RepRap model Rebelix [2] at our department. Its construction enables precise movement of the instrument in three dimensions. The extruder of the filament represents the instrument in this case.

We had the idea, that we have similar problem, but we need to move only in two dimensions – left - right and up - down.

So, we decided, that we base the project solution on this 3D printer construction. This was not a problem because the Rebelix printer is designed as the open hardware. It also means that the list of the parts used is available on the Internet.

We prepared the initial scheme of a scanner and ordered all the needed parts. As the control board we have chosen the Mini-Rambo [3] board, the new version 1.3 (Figure 2). Its

delivery from the USA was the longest waiting period during the construction.

The Mini-Rambo is as a matter of fact Arduino Mega with included drivers for the five motors. So, the advantage of this board is the possibility to use Arduino IDE to program it. First, we thought, that we would write completely new program to

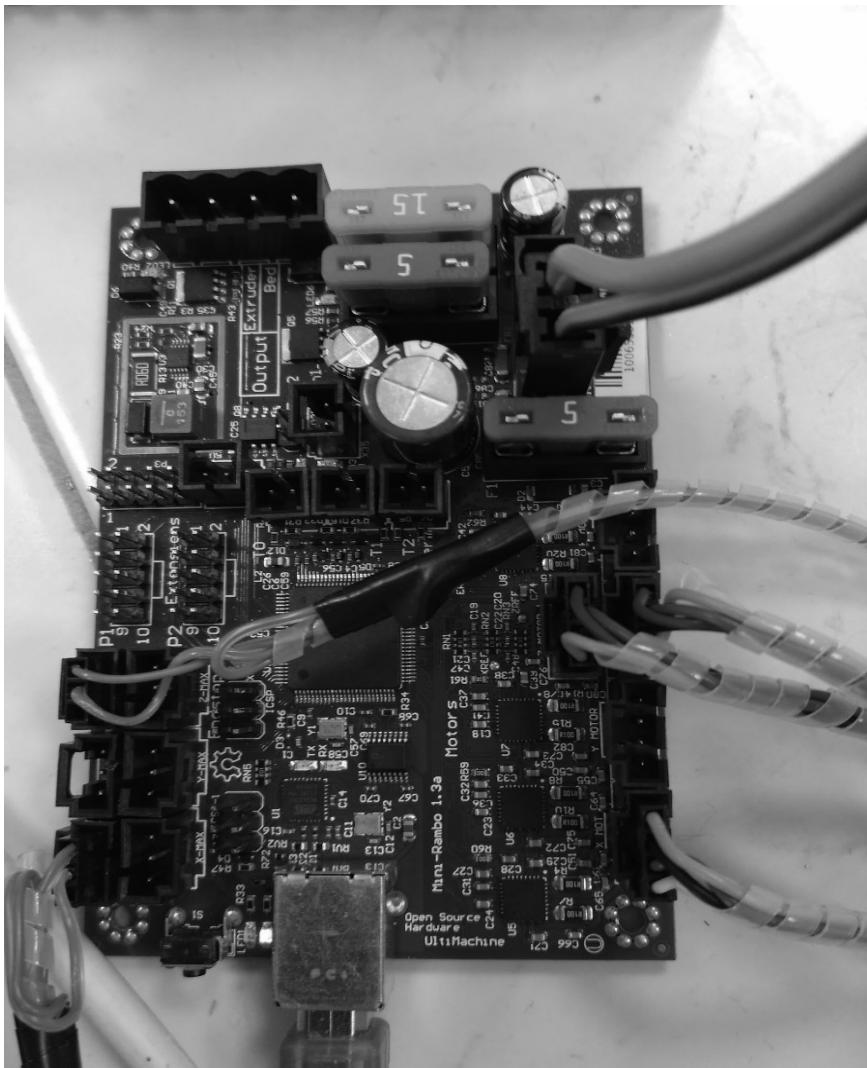


Fig. 2. Control board Mini-Rambo 1.3.

control our scanner. Later we decided to use the original firmware, which is used to control 3D printers. Again, it was not a problem, because it is open-source software.

This means, that we control this board by the G-Code<sup>1</sup> commands. We needed leading computer, which would send the G-Code commands to the board.

#### IV. CONSTRUCTION

After arrival of all the parts (except the board) we began the construction. With the help of our technicians we prepared stake, threads and sockets. In comparison to construction of

3D printer 3 years ago, we had simpler task because there are motors with connected screw-thread available now, so it is not needed to join screw-thread to motor.

Next, we needed the parts printed by 3D printer. We used some parts directly from 3D printer, but we also needed to prepare and print some new parts. We made them by the online support CAD software Tinkercad [6]. Then we printed all the parts on our printer.

Then it was necessary to write a code to send desired G-Code to the control board. It was required to cover the entire frame with gel by laser ray by moving the scanner. So, we decided to use Lissajous curve for the movement of the frame.

---

<sup>1</sup> G-Code is the language for the digitally controlled machines. I will show an example of the G-Code later.

As the programming language we selected Python for its simplicity and good portability and as the operating system for the control computer we used Linux. The computer is connected by USB cable type A.

The advantage of the presented solution was that we could test the first versions of our program on the 3D printer. It was important because we had all construction complete except for the board, which arrived two months later.

Finally, three months later, we had the construction complete (Figure 3 and 4) and we could test it. The meanings of the items in the figure 3 are:

1. Stepper motor for the movement of the frame in the x axis direction.
2. Stepper motor for the movement of the frame in the z axis direction.
3. Coulisse for the screw-threads.
4. Screw-threads.
5. Leading threads.
6. Coulisse for the leading threads.
7. Gear transmission for the movement of the frame.
8. Frame for the holding of the sample.
9. Aluminum profile for the stability of construction.

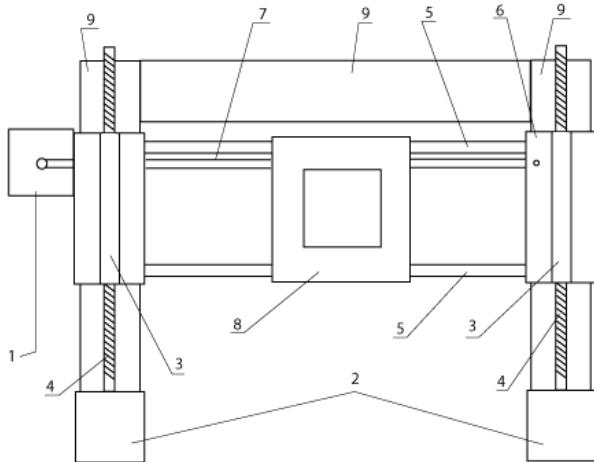


Fig. 3. Scheme of the scanner.

## V. TESTING AND TUNING

All the problems, which we solved during testing the scanner, are described in this chapter.

First big problem was the selection of the correct firmware for the Mini-Rambo board. First, we tried the Marlin firmware, but we had a problem with end-stops<sup>2</sup>. So, we tried alternate firmware RAMPS, but it was even worse. So, we used the development version of the Marlin finally and we had to make some changes in the source code. For example, we needed to disable setting for unused y axis (forwards – backwards direction). We had to change level of the current for motors because with the original settings the motors are overheated.

<sup>2</sup> End-stops are the tools which prevent the moving parts from getting out of the leading bars

We had to test the developed program, which generated G-Code, carefully. First, we used relative coordinates but finally we used absolute coordinates to eliminate rounding induced errors. Here is an example of the source code and example of the generated G-code.

## VI. EXAMPLE OF THE SOURCE CODE

```
import serial
ramecek =
serial.Serial ("/dev/ttyACM0",baudrate=115
200,
timeout=0)
ramecek.write("G21\n".encode())
XS = 80
ZS= 40
stred="G0 X"+str(XS)+" Z"+str(ZS)+"\n"
ramecek.write("G90\n".encode())
ramecek.write("M92 X100 Z390\n".encode())
ramecek.write("G28 X Z\n".encode())
ramecek.write(stred.encode())
```

This code sets up a communication between computer and control board and then sends G-Code commands via USB cable. The whole program is much bigger and complicated and has about 320 lines of the code.

## VI. EXAMPLE OF THE G-CODE

And here is the example of the G-Code for the moving frame of the scanner:

```
G0 X71.601 Z32.191
G0 X73.018 Z31.602
G0 X74.09 Z30.734
G0 X74.694 Z29.738
```

These are the commands for the movement of the frame along the x axis (right - left direction) and along z axis (up - down direction). For the matter of the interest there are about 2000 G-Code commands for one complete covering of the frame with Lissajous curve.

## VII. PROBLEMS AND CALIBRATION

The biggest problem presents the requirement to make small pauses between commands. It is not possible to send a command after a command without the pauses because the control computer generates the commands faster than the scanner can perform them. The duration of the pauses depends on four variables – two dimensions of the sample and two parameters of the Lissajous curve. If the pauses are too small, the control board can't correctly process G-Code to the scanner in time or the scanner can't execute (more precisely finish) all the commands. In opposite, if the pause is too big, the movement is abrupt and then the scanner shakes.

So, we have made the algorithms in which the system is learning itself. First, we estimate the duration of the pauses according to all enumerated parameters and then the

calibration starts. Control computer tries to work with estimated pauses and reads answers from the control board. If there is any error, the system stops for a while and prolongs the pause. First calibration ends when the system passes all the

commands without any error five times. When there appears no error during all the calibration, the pauses are cut by 30 % and calibration repeats.

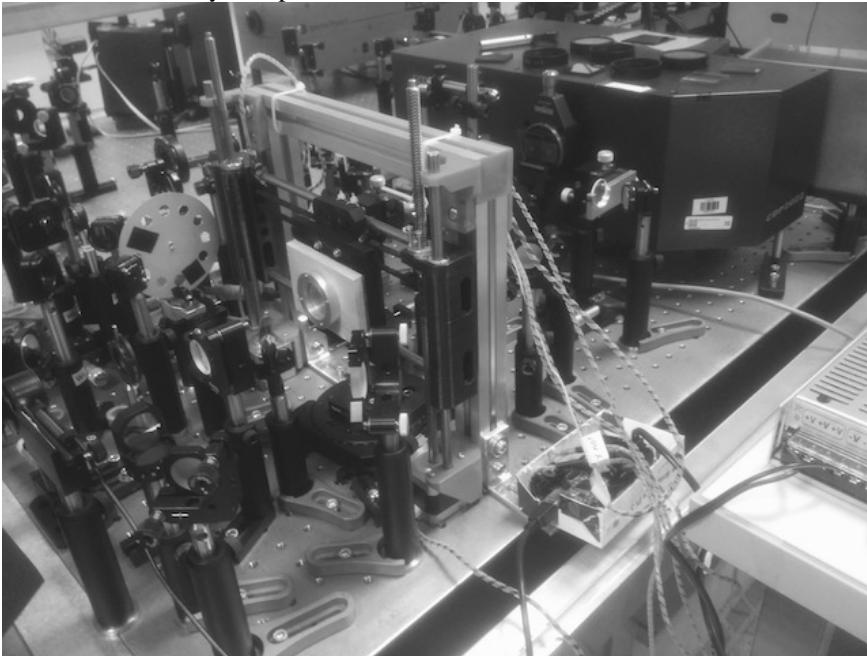


Fig. 4. Final form of the scanner.

After this first calibration or calibrations there comes the second calibration. The whole program runs once more with the same conditions as during the real use later. The system watches for the errors again and if any appear, the appropriate pause is prolonged. Again, the process ends when no error appears five times.

After second calibration computer notifies the user that everything is ready to run, and user can start the lasers.

During all the development we worked on the normal PC, but we prepared one-board computer Raspberry Pi 2 with attached touch display as the control computer for the real use.

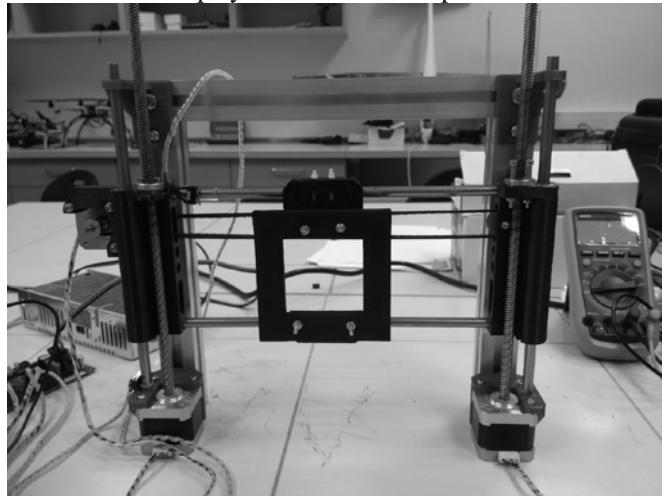


Fig. 5. The scanner on the working position.

### VIII. CONCLUSION

We managed to make the required scanner, which was named the Lissajouze scanner. The submitter is satisfied with our work and scanner will serve to the research. The system

partly fulfils the standards of the Industry 4.0 because it is self-learning. It would have to read the dimension and the position of the sample itself to meet the standards completely.

The Lissajouze scanner is used for moving of this kind of the sample where is necessary to excite the researched molecules so that no previous history of the sample (e.g. illumination) has any meaning. (e.g. the gel situated in the test-glass) The excited molecules in the sample are researched with the methods called pump-and-probe, which is common in the scope femtosecond optical spectroscopy.

We got very valuable experiences during this project and these experiences can be used in following constructions. We will be very pleased if it helps anybody to make any similar construction.

### REFERENCES

- [1] Institute of Physics and Biophysics [online]. České Budějovice, 2017 [cit. 2017-04-03]. Available from: <http://www.prf.jcu.cz/en/ufy/>
- [2] Marlin - RepRapWiki [online]. 2016 [cit. 2017-04-03]. Available from: <http://reprap.org/wiki/Marlin>
- [3] Mini-Rambo 1.3 - Ultimachine [online]. 2017 [cit. 2017-04-03]. Available from: <https://ultimachine.com/products/mini-rambo-1-3>
- [4] Raspberry Pi - Teach, Learn and Make with Raspberry Pi [online]. Raspberry Pi Foundation, 2017 [cit. 2017-04-03]. Available from: <https://www.raspberrypi.org/>
- [5] Rebelix - 3D tiskárna [online]. Praha, 2017 [cit. 2017-04-03]. Available from: <http://reprap4u.cz/rebelix/>
- [6] Tinkercad: Create 3D digital designs with online CAD [online]. Autodesk, 2017 [cit. 2017-04-04]. Available from: <https://www.tinkercad.com>

# Methods of Crypto Protection of Color Image Pixels in Different Code Systems

Nataliia Vozna<sup>1</sup>, Yaroslav Nykolaichuk<sup>1</sup>, Orest Volynskyi<sup>2</sup>, Petro Humennyi<sup>1</sup>, Andrij Sydor<sup>1</sup>

1. Department of Specialized Computer Systems, Ternopil National Economic University, UKRAINE, Ternopil, 8 Chekhova str., email: [nvozna@ukr.net](mailto:nvozna@ukr.net)

2. Department of Cyber Security, Ternopil National Economic University, UKRAINE, Ternopil, 8 Chekhova str., email: [orestsks@ukr.net](mailto:orestsks@ukr.net)

**Abstract:** The relevance of the development of theoretical foundations, methods and algorithms for encoding color image pixels by the problem-oriented multifunctional data structuring and the representation of color image code pixels in Rademacher (R), Krestenson (K), Rademacher-Krestenson (RK), Haar-Krestenson (HK) and Galois (G) Systems is substantiated in this article. The purpose of the research is to increase the efficiency of the algorithms for digital image transforms, processing and recognition using modular arithmetic of extended Galois fields on the basis of mathematics of arithmetic operations of a non-positional residue number system.

**Keywords:** crypto protection, color image pixels, Rademacher and Krestenson Systems, Residue Number System.

## I. INTRODUCTION

Successful development of modern computer technology, microelectronics and telecommunication systems promotes designing and mass production of color TV displays as well as personal computers, mobile devices, camcorders, tablet PC screens, industrial and large format color displays.

The large-scale application of various types of video equipment in all branches of industry and their wide-spread personal use determines a high level of importance of the solutions to theoretical and applied problems of increasing and optimizing the efficiency of video image structuring during the processes of creation, encoding, transformation, crypto protection, transmission, archiving and access receiving to color images as well as their use.

The examples of setting and successful solving the problems referring to this issue on the basis of the mathematical foundations development, the implementation of the algorithms and hardware and software tools for image processing and recognition were thoroughly highlighted in the works of scientific researches [1-5].

Considerable attention is paid to solving research problems in this field and creating algorithms of the image structural properties and features.

## II. METHODS OF MULTIFUNCTIONAL STRUCTURING OF COLOR IMAGE PIXELS IN THE SYSTEM OF EXTENDED GALOIS FIELDS

The analysis of the mathematical foundations of the existing algorithms for color image processing and

recognition was carried out by segmentation methods on the basis of histogram thresholding and cumulative histograms. It is analysis of the statistic estimates of the mean value, dispersion, asymmetry and the degree of contrast of the intensity histograms homogeneity taking into account the dispersion of pixels coordinates of image fragments and silhouettes, as well as image clustering methods [1-5].

As a result, it was found that the main components of the algorithms of the above-mentioned methods for image processing are the following arithmetic operations: summarizing ( $\sum x_i$ ), division ( $P(i) = n_i / n_0$ ), absolute difference ( $|x_i - x_j|$ ), square ( $x_i^2$ ), multiplication ( $x_i \times x_j$ ), square difference ( $[x_i - x_j]^2$ ), sum of multiplication ( $\sum x_i x_j$ ), which are commonly performed due to the low-speed arithmetic of the binary number system.

## III. THE METHOD FOR ENCODING RGB PIXELS IN THE RADEMACHER AND KRESTENSON SYSTEMS

According to the international RGB color model, colors are presented as a combination of three main colors: red (R), green (G) and blue (B) [4].

In this case in the computer RGB system, the main color has 256 gradations. Thus, the color code of the RGB system is made up of three bytes, that is, 24 bits in the Rademacher system.

The colors of the Hamming distance pixels on a monitor, given in Cartesian coordinates, can be coded in the Residue Number System (K). This is implemented by introducing three relatively simple modules ( $P_1, P_2, P_3$ ), which allow encoding each pixel of the RGB system in the binary system by forward integer transform of the residue number system (RNS) according to the expression [6]:

$$N_k = \text{res} \sum_{i=1}^3 b_i \cdot B_i \pmod{P_0} \quad (1)$$

where  $B_i$  - the orthogonal bases of RNS, which are calculated according to diophantine equations:

$$B_1 = P_2 \cdot P_3 \cdot m_1 \equiv 1 \pmod{P_1}; \quad (2)$$

$$B_2 = P_1 \cdot P_3 \cdot m_2 \equiv 1 \pmod{P_2}; \quad (3)$$

$$B_3 = P_1 \cdot P_2 \cdot m_3 \equiv 1 \pmod{P_3}, \quad (4)$$

where  $m_1, m_2, m_3$  - inverse elements of the RNS [8];

$P_0 = P_1 \cdot P_2 \cdot P_3$  - color image pixel encoding range with color depth  $K_0 = \hat{E}[\log_2 P_0]$ ,  $\hat{E}[\bullet]$  - integer function with rounding to a larger integer.

RGB pixels encoding in the Rademacher-Krestenson system is provided by selecting the following values of the encoding range of  $b_i$  remainders in the Rademacher system:

$$b_1 = b_R; \quad 0 \leq b_R \leq 255; (00000000 \div 11111111);$$

$$b_2 = b_G; \quad 0 \leq b_G \leq 255; (00000000 \div 11111111);$$

$$b_3 = b_B; \quad 0 \leq b_B \leq 255; (00000000 \div 11111111).$$

In addition, taking into account the coefficients  $m = 1.0$ ,  $n = 4.5907$ ,  $p = 0.0601$ , in order to achieve the most saturated green color, the range of its change can be set as  $0 \leq b_G \leq 254$  that provides relevant simplicity of the following modules:  $P_1 = 256$ ,  $P_2 = 255$ ,  $P_3 = 257$ .

To verify the relevant simplicity of the selected modules system, they are factorized into multipliers:  $256 = 2^8$ ,  $255 = 5 * 51$ ,  $257$  - a prime number, i.e.  $P_0 = 16776960$ , where  $P_0 < 2^{24} = 16777216$ . That is, the condition for creating a 24-bit pixel code in the Rademacher-Krestenson System is satisfied.

In binary system module codes are represented as:

$$P_1 = 100000000_{(2)}, P_2 = 11111111_{(2)}, P_3 = 100000001_{(2)}.$$

$$\text{Then: } P_0 = 11111111111111100000001_{(2)}.$$

As a module  $P_1 = 2^8$  is among the modules  $P_1, P_2, P_3$ , then, according to the inverse RNS transform, the remainder of  $N_k$  (G - color features) will be presented without decoding it by eight low orders of  $N_k$ , which is in the Rademacher system.

According to the Diophantine equations solution (2-4), the following values of the inverse elements  $m_i$  and basic numbers  $B_i$  are received:

$$m_1 = 255, \quad B_1 = 16711425; \quad m_2 = 128, \quad B_2 = 8421376; \\ m_3 = 129, \quad B_3 = 8421120$$

The verification of the calculation accuracy of the RNS transform is performed according to the equation:

$$N_k = (b_R \cdot B_1 + b_G \cdot B_2 + b_B \cdot B_3) \cdot (\text{mod } P_0) = 1 \quad \text{when} \\ b_R = 1, \quad b_G = 1, \quad b_B = 1.$$

That is,

$$N_k = (1 \cdot 16711425 + 1 \cdot 8421376 + 1 \cdot 8421120) \cdot (\text{mod } P_0) = 1.$$

For example,  $R = 10$ ,  $G = 200$ ,  $B = 100$ .

Then

$$N_k = (10 \cdot 16711425 + 200 \cdot 8421376 + 100 \cdot 8421120) \cdot \\ \cdot (\text{mod } 16776960) = 9187850$$

which corresponds to the binary representation of the RGB pixel in the Krestenson System  $(100011000011001000001010)_2$ .

Decoding of such representation is as follows:

$$r_i = \text{res}N_k \pmod{P_1}; \quad g_i = \text{res}N_k \pmod{P_2}; \\ b_i = \text{res}N_k \pmod{P_3}.$$

#### IV. THE METHOD FOR COLOR IMAGE PIXELS ENCODING IN THE RADEMACHER-KRESTENSON AND THE HAAR-KRESTENSON SYSTEMS

The encoding of color image pixels according to the RGB color model is carried out by the 24-bit binary code, when the intensity of each of the colors is represented by the 8-bit binary code of the Rademacher System:

$$R \begin{cases} r_{8-1} \\ \dots \\ r_i \\ \dots \\ r_0 \end{cases}; \quad G \begin{cases} g_{8-1} \\ \dots \\ g_i \\ \dots \\ g_0 \end{cases}; \quad B \begin{cases} b_{8-1} \\ \dots \\ b_i \\ \dots \\ b_0 \end{cases}$$

$$0 \leq r_i \leq 255; \quad 0 \leq g_i \leq 255; \quad 0 \leq b_i \leq 255.$$

Encoding of the color image RGB pixels in the Rademacher-Krestenson (RK) and Haar-Krestenson (HK) Systems is carried out by selecting relatively simple modules system  $(P_1, P_2, P_3)$ , whose product exceeds the range of quantization of the brightness values  $(r_i, g_i, b_i)$ .

Such a condition can be satisfied by a different set of the RNS discrete transformer modules, for example,  $P_1 = 5, P_2 = 7, P_3 = 8$ , which provide encoding of  $r_i, g_i$  and  $b_i$  brightness in  $P_0 = 5 * 7 * 8 = 280 > 255$  range. The following code structure is created in the R-K System, which unambiguously represents the corresponding RGB-pixel code:

$$R \vee G \vee B \begin{cases} a_2 \\ a_1 \\ a_0 \end{cases}; \quad \begin{cases} c_2 \\ c_1 \\ c_0 \end{cases}; \quad \begin{cases} d_2 \\ d_1 \\ d_0 \end{cases}$$

$$P_1 = 5 \quad P_2 = 7; \quad P_3 = 8,$$

where  $a_i \in \overline{0,1}$ ;  $c_i \in \overline{0,1}$ ;  $d_i \in \overline{0,1}$ ;  $i \in \overline{0,2}$ .

In this case, each value  $a_i, c_i, d_i$  is calculated as the remainder according to the expressions:  $a_i = \text{res}(r_i \bmod P_1)$ ;  $c_i = \text{res}(g_i \bmod P_2)$ ,  $d_i = \text{res}(b_i \bmod P_3)$ .

For a given set of modules, the inverse elements  $m_i$  and the basic numbers  $B_i$  are determined according to the Diophantine equations solutions (2-4):

$$m_1 = 1, \quad B_1 = 56, \quad m_2 = 3, \quad B_2 = 120, \quad m_3 = 3, \\ B_3 = 105.$$

Accuracy of the obtained  $m_i$  and  $B_i$  values is verified according to the expression (1):

$$N_1 = (1 \cdot 56 + 1 \cdot 120 + 1 \cdot 105) \bmod 280 = 1$$

For example, the following values of color intensity of the RGB-pixel are set as:  $r_i = 10$ ,  $g_i = 100$ ,  $b_i = 37$ .

Then, RGB-pixel codes are received in the Rademacher System:

$$r_i = 00001010_{(2)}; \quad g_i = 01100100_{(2)}; \quad b_i = 00100101_{(2)};$$

in the Rademacher-Krestenson system:

$$r_i = (\overbrace{000011101}^{P_1} \overbrace{00010010}^{P_2} \overbrace{0}^{P_3})_{(5,7,8)}; g_i = (\overbrace{000010010}^{P_1} \overbrace{00010010}^{P_2} \overbrace{0}^{P_3})_{(5,7,8)};$$

$$b_i = (\overbrace{010010101}^{P_1} \overbrace{00010010}^{P_2} \overbrace{0}^{P_3})_{(5,7,8)}.$$

Representation of the RGB pixel code for each  $r_i$ ,  $g_i$  and  $b_i$  intensity value in the Haar-Krestenson System is made according to the structure:

$$R \vee G \vee B \begin{cases} a_{P_1-1} \\ \dots \\ a_i \\ \dots \\ a_0 \end{cases}; \quad \begin{cases} c_{P_2-1} \\ \dots \\ c_i \\ \dots \\ c_0 \end{cases}; \quad \begin{cases} d_{P_3-1} \\ \dots \\ d_i \\ \dots \\ d_0 \end{cases}$$

$$P_1 = 5 \quad P_2 = 7; \quad P_3 = 8,$$

where  $i \in \overline{0, P_i - 1}$

For the specified color intensity values of the RGB pixel  $r_i = 10$ ,  $g_i = 100$ ,  $b_i = 37$ , the following code structure in the H-K system is obtained:

$$r_i = (10000..0001000..00000100);$$

$$g_i = (10000..0010000..00100000);$$

$$b_i = (00100..0010000..00000100).$$

The representation of  $r_i$ ,  $g_i$  and  $b_i$  color brightness digital values in different systems leads, correspondingly, to different code length according to the expressions:

1.  $K_R = \log_2 2^8 = 8$  bits in the Rademacher System (R).

2.  $K_{R-C} = \sum_{i=1}^3 [\hat{E}(\log_2 P_i - 1)] = 3 + 3 + 3 = 9$  bits in the Rademacher-Krestenson system (R-K).

3.  $K_{H-C} = \sum_{i=1}^n P_i = 5 + 7 + 8 = 20$  bits in the Haar-Krestenson System (H-K).

## V. STRUCTURE DEVELOPMENT AND EXPERIMENTAL STUDIES OF STRUCTURAL, TIME AND HARDWARE COMPLEXITY OF ADC WITH THE R AND H-K OUTPUT CODES.

It is expedient to make multifunctional encoding of RGB pixels in the R-K and H-K systems at the level of analog-to-digital conversion of the analog signals intensity of the RGB sensors. Such a principle of multifunctional data structuring in color formation is implemented by parallel ADC, the structure of which is shown in Fig 1.

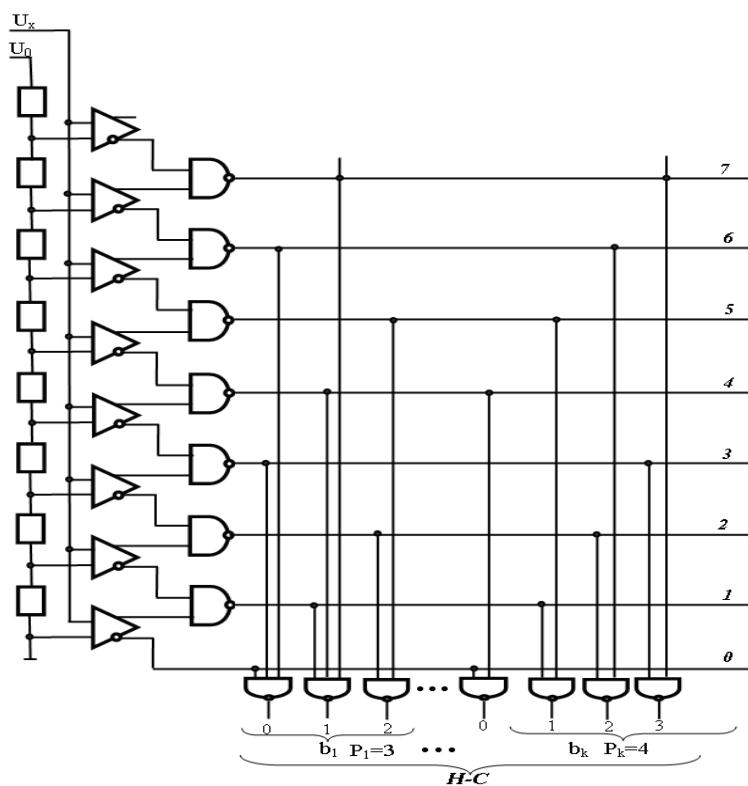


Fig.1. The structure of a multi-purpose parallel ADC with output codes in the Haar-Krestenson System.

ADC consists of 1 – input analogue bus; 2 – paraphase comparators; 3 – input reference bus; 4 – exemplary resistors; 5 – the first logic elements "AND-NOT"; 6 – the second logic elements "AND-NOT", 7 – output ADC bus.

ADC efficiency is determined according to the expression:

$$\tau_{ADC_2} = \tau_{k_2} + \tau_{LE_2} + \tau_{LE_3},$$

where  $\tau_{k_2} = 2\nu$  - switching time for paraphase comparator;

$\tau_{LE_2} = 1\nu$  - switching time for two-input logic element "AND-NOT";

$\tau_{LE_3} = 1\nu$  - switching time for multi-input logic element (LE) "AND-NOT";

That is, the efficiency of ADC is determined by the total delay of signals:

$$\tau_{ADC_2} = (2+1+1)\nu = 4 \text{ micro cycles.}$$

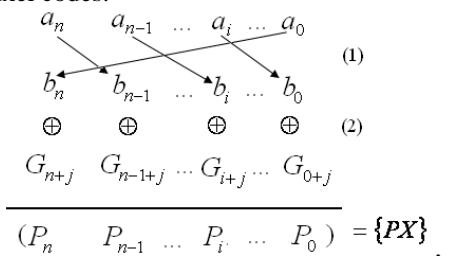
When calculating the time complexity of the ADC components, it is taken into account that the switching time of the paraphase comparator is 2.5 times less in comparison with the single-phase comparator due to positive trigger feedback between the direct and inverse outputs.

## VI. THE METHOD OF CRYPTO PROTECTION OF COLOR IMAGE RGB PIXELS.

Crypto protection of the RGB image pixels is performed in order to restrict unauthorized access to color images that are generated in real time. It's encoded in different number systems, transmitted via communication channels, recorded in database storage, and displayed on the user monitors. There are different methods for encrypting files containing color image data and data arrays, which include a certain amount of color images. In this case, information systems use standard algorithms for data arrays protection from unauthorized access on the basis of hashing, symmetric and asymmetric RSA algorithms, elliptic curves, etc. [7, 8].

The method for encryption of color images RGB pixels, which are represented by R, R-K and H-K codes of the described methods, is proposed. In this case, structured R-K and H-K codes are problem-oriented to increasing the efficiency of the image transform, processing and recognition in accordance with the modular arithmetic of the Residue Number System.

It is expedient to apply an effective method based on hashing of certain code positions and logic combination of bits of generated Galois sequences [9] according to the following graphs as the main method of crypto protection of RGB pixel codes:



where  $a_i$  - bits of R-K or H-K pixel codes; 1 - hashing procedure ( $b_i := b_j, i \neq j, i \in \overline{0, n}$ ),  $P_i, i \in \overline{0, n}$  - created code of crypto protected pixel  $PX$ .

Bits of Galois  $\{G_i\}$  codes are generated according to secret keys.

## VII. CONCLUSIONS

The relevance of the development of the theory, methods and algorithms for encoding color image pixels and their representation in different systems has been

substantiated. This allows to increase the efficiency of algorithms for digital image transform, processing and recognition on the basis of the mathematics of arithmetic operations of the non-positional Residue Number System.

The analysis of the mathematical foundations of existing algorithms for color image processing and recognition was carried out by segmentation methods on the basis of histogram thresholding and cumulative histograms, statistic estimates of the mean value, dispersion, asymmetry and the degree of contrast of the intensity of histograms. This is exemplified by homogeneity taking into account the dispersion of pixels coordinates of image fragments and silhouettes, as well as image clustering methods.

It is proposed to carry out structured encoding of color image pixels by the codes of non-positional number systems of R-K, H-K and G. This allows to increase the efficiency of algorithms for image processing by 2-3 orders.

## REFERENCES

- [1] Otsu N., A threshold selection method from grey level histograms, *IEEE Trans. Systems Man Cybernet*, No.9, pp.62-66, 1979..
- [2] Zhang Yudong and Wu Lenan., Fast Document Image Binarization Based on an Improved Adaptive Otsu's Method and Destination Word Accumulation, *Jurnal of Computational Information Systems*, No.6, pp.1886-1892, 2011.
- [3] U. Ramer, "An Iterative Procedure for the Polygonal Approximation of Plane Curves," *Computer Graphics Image Processing*, Vol. 1, No. 3, pp. 244-256, 1972.
- [4] R. Melnyk Algorithms and methods for image processing: Teaching manual., Lviv: Lviv Politechnika Publishing House, 220 pp., 2017.
- [5] N. Lotoshynska. Theory of color and color formation: Teaching manual, Lviv: Lviv Politechnika Publishing House, 204p., 2014.
- [6] N. Vozna., Y. Nykolaichuk and N. Shyrmovska, "Method of formation of structured data of quasi-stationary objects on the basis of the Residue Number System of the Krestenson basis", *Scientific and Technical Journal "Exploration and Development of Oil and Gas Fields*, No. 3 (40), pp.62-65, 2011.
- [7] Ya. Nykolaychuk, M. Kasianchuk and I.Yakymenko, "Theoretical Foundations for the Analytical Computation of Coefficients of Basic Numbers of Krestenson's Transformation", *Cybernetics and Systems Analysis*, Volume 50, Issue 5, pp. 649-654, September, 2014.
- [8] Ya. Nykolaychuk, M. Kasianchuk and I.Yakymenko, "Theoretical Foundations of the Modified Perfect form of Residue Number System", *Cybernetics and Systems Analysis*, Volume 52, Issue 2, pp. 219-223, March, 2016.
- [9] Y. Nykolaichuk, Galois Field Codes: Theory and Application, Ternopil: Ltd.: Terno-graf, 576 pp., 2012.

# Smart Information Gathering Support of Mechatronic System

Orest Ivakhiv, Markian Nakonechnyi, Oleksandr Viter, Radyslav Behota

Computer Technology, Automation and Metrology Institute, Lviv Polytechnic National University, UKRAINE, Lviv, 12 Bandera str., email: [oresti@poly.net.lviv.ua](mailto:oresti@poly.net.lviv.ua)

**Abstract:** The smart programmable systems often are very convenient for the measurement object semantic reducing. The object state analyzing unit is the core of the system. There is considered that an object behavior a priory is unknown. Corresponding to this mode it is necessary to observe an object state and prepare a proper survey program. The main characteristics of proposed system are investigated in this paper.

**Keywords:** smart, programmable, flexible, adaptive, activity, compression, system.

## I. INTRODUCTION

A mechatronic system involves a set of sensors, conditioners of sensors' signals, processing and decision-making unit and actuator [1]. Combination system software and hardware with Internet as a communication link give us a cyber-physical system [2]. Obviously, information-gathering support is realized as multiplex time division systems. They are widely shared in the different areas of human activity such as scientific research, aeronautics, agriculture, space investigation, image processing and data monitoring systems for health care, which simultaneously monitors, transmits by radio a records data relating to a plurality of physiological parameters etc. [3-6]. Apart from the demands for small size, lightweight and long operational lifetime, the sensor systems should preferably also be flexible, versatile and intelligent [6]. The traditional approach of reconstructing signals or images from measured data follows the well-known Shannon sampling theorem, which states that the sampling rate must be twice the highest frequency. It was found that at any given time, only a fraction of the neurons were active therefore it was possible to reduce data by 97%. Therefore, it is necessary to use some compression techniques possibility.

The aim of data compression is to reduce redundancy in stored or communicated data, thus increasing effective data density. Data transmission, compression and decompression of data files for storage is essentially the same task as sending and receiving compressed data over a communication channel. Compressive sensing is a new type of sampling theory, which predicts that sparse signals and images can be reconstructed from what was previously believed to be incomplete information [7,8].

Sensors set which are checking behaviors of object parameters reflect an object state. Sampling interval of a sensor signal is corresponded with both signal frequency properties and desired error of analog signal renovation at the receiver side. It was taken partly stationary zero-mean random process as every signal mathematical model. Any stationary interval differs from another by the frequency

properties, which are corresponded with the form of correlation function or its parameters values. These properties are discovered by every the  $i$ -th source activity manifestation. A proper regular type system-sampling program depends on this set of a priory known or estimated activities. If the object state a priory is poor known, it needs adaptation to the current object situation. It is typical remote investigated objects, i.e. the deep space invention instrumentation or dirty territory serving. In this case, the intelligent multiplex system implantation is considered as well operating and the intelligent measurement instrumentation functions should be extended by the implantation of the task of the observed object state identification, the inspected parameters real activities to the adopted sampling program adequacy learning and in this sense the external situation registration [9].

## II. REGULAR TYPE SYSTEM STRUCTURE

The structure of the smart multi-channel tool (Fig. 1) includes a unit of a measurement object totality sources behavior analyzing BAU [10], an object state observing unit SOU, a unit of survey programs storage PSU, a unit CMU combining codec/modem functions, i.e., an analog-to-digital conversion, a noise immunity encoding and modulation, which is connected to the communication link (point1).

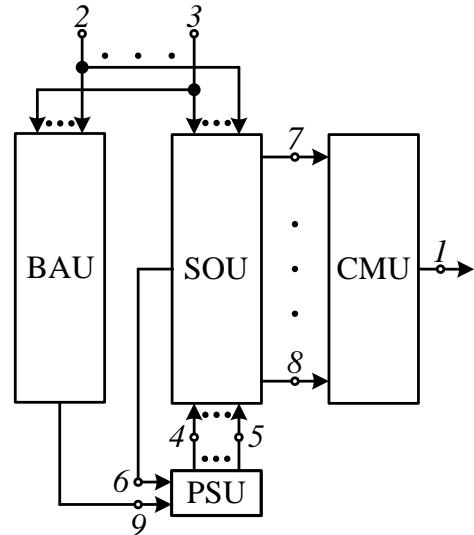


Fig.1. The structure of the multi-channel tool

Outputs of analog signal sources ( $i = \overline{1, n}$ ) simultaneously are connected to the corresponding inputs of a unit BAU (points 2 and 3) and an observing unit SOU. A unit PSU sets for the unit SOU the sequence of source sampling procedure (points 4 and 5). A unit SOU observes the accumulation of

absolute sampling errors from the sources totality and after analyzing procedure, if it is necessary informs the program storage unit PSU about the need to change the survey program using the signal at the point 6. Simultaneously, a unit BAU analyzes analog sources current activities.

## II. OBJECT BEHAVIOR STATE ESTIMATION UNIT

For characteristics setting of this unit a compliance of survey software to the current situation can be accessed through tracking amount behavior of the sampling errors from all totality sources during a cycle of the survey. As well as there are added the random variables therefore relevant point and interval statistic estimates (thresholds) can be found for their sums. No exceeding of this threshold with a given confidence probability is identical to matching the real current situation at the measurement object, and its exceeding it is the message about the current situation change and demands on the need of transition to a more relevant survey program. Comparison of the threshold value and the sum of sampling errors is carried out at every step of the survey.

In this unit (Fig.2) outputs of analog sources are connected to the proper inputs (points 2 and 3) of unit SOU.

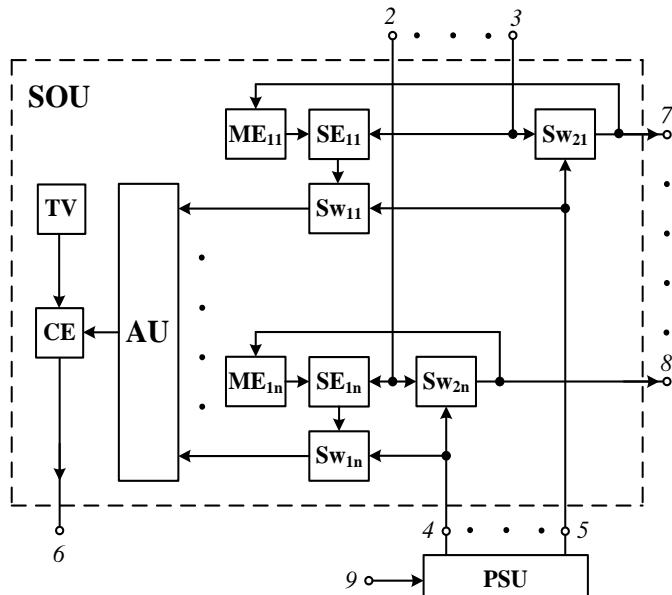


Fig.2.The structure of object state observing unit SOU

At the very beginning (at the first step) the sample values from each source ( $i=1, n$ ) are recorded in its memory element  $ME_{1i}$  through the open switch  $Sw_{2i}$ . At the second and subsequent steps the sample value of each source enters to the subtraction element  $SE_{1i}$  and passes through open switch  $Sw_{2i}$  and the Codec/Modem unit CMU to the device output, i.e., to the communication link (point 1) corresponding with a proper sampling program. The switch  $Sw_{2i}$  is open by the control signal from programmer storage PSU (points 4 and 5), if it is provided by the current sampling (survey) program. Meantime this signal closes the switch  $Sw_{1i}$  for the same  $i$ -th channel. Therefore, the difference from the  $i$ -th subtraction element  $SE_{1i}$  is disconnected from an adder unit AU, i.e., it does not take part in totality sources sum formation. In the element  $SE_{1i}$  the value recorded in the  $ME_{1i}$  at the previous step is subtracted from the current

sampled value at the each sampling tact. On the next step, the subtrahend will be this new value, and not the first sample value of the  $i$ -th source

Let us consider that sensor analog signals are well described by mathematical model as zero-mean partly stationary random process. So, let us take some realization  $u_i(t)$  of the  $i$ -th random signal (Fig.3) which is regular sampled in moments  $t_j$  and  $t_{j+1}$  with a sampling interval equal to  $T_0 = t_{j+1} - t_j$ . In this case, one can obtain an absolute and its relative value is as follow

$$\overline{\Delta u_{si}^2} = \frac{1}{3}(\omega_{li} \cdot \sigma_i \cdot T_{oi}) \text{ or } \delta_{si}^2 = \frac{\overline{\Delta u_{si}^2}}{\sigma_i^2} = \frac{1}{3}(\omega_{li} \cdot T_{oi})^2, \quad (1)$$

here  $\omega_{li}$  and  $\sigma_i$  are mean-square frequency and mean-square deviation of the  $i$ -th signal, and thus,  $T_{oi}$  is its sampling interval ( $\omega_{li}^2 = \int_0^\infty \omega^2 G(\omega) d\omega / \int_0^\infty G(\omega) d\omega$ , here  $G(\omega)$  is power spectral density).

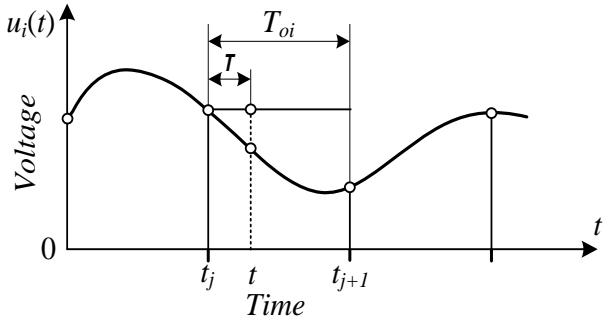


Fig.3. Sampling error vs time dependence ( $\tau = t - t_j$ )

If a sampling program coincides with an object current state then a sum of sampling error from all sources totality (at the output of adder unit AU) do not exceed the settled threshold value (of the output of unit TV). But when an object state changes due to its new environment situation then used sampling program becomes wrong, everyone source sampling error and its sum also become differ from supposed for same partly stationary interval. It demands to change sampling program and therefore corresponding signal appears at the output of comparative element CE (point 6).

Compliance of survey program to the current situation is determined according to mean-square deviation of summarized sampling error from all sources together. In accordance with the law of large numbers, it can be assumed that the total error as a random value will be well described by Student's or Gauss's distribution law. Therefore, with some credential probability  $P_{tol}$  one can set the guarantee interval for the total sampling error

$$x_{tol} = \pm t_\alpha \cdot \sigma_{s\Sigma} \text{ or } X_{tol} = \pm t_\alpha \cdot \bar{\sigma}_{s\Sigma}, \text{ and } \delta_{s\Sigma} = \sqrt{\sum_{i=1}^n \delta_{si}^2}, \quad (2)$$

here,  $t_\alpha$  is the guarantee coefficient of the set credential probability. The resulting error from the output of AU is compared with the calculated by unit SOU threshold value (2) from TV element. Signal of excess is fed into the PSU unit (point 9).

Since the sampling error is sign-alternating, its average value is equal to zero and the variance coincides with the second raw moment. Therefore if errors of all channels of a multi-channel device are independent, total error variance is estimated as follows

$$\overline{\sigma_{s\Sigma}^2} = \sum_{i=1}^n \Delta u_{si}^2 = \frac{1}{3} (\omega_{li} \cdot \sigma_i \cdot T_{oi})^2 \text{ or } \delta_{s\Sigma}^2 = \frac{1}{3} \sum_{i=1}^n (\omega_{li} \cdot T_{oi})^2 \quad (3)$$

Note that since the power of the  $i$ -th measurement signal is described by expression  $\sigma_i^2 = \frac{1}{2\pi} \int_0^\infty G_i(\omega) d\omega$  and generalized spectral power density  $G_\Sigma(\omega) = \sum_i G_i(\omega)$ , certain individual signals being independent, the following equality will be true:  $\sigma_\Sigma^2 = \sum_i \sigma_i^2$

### III. ANALYZING UNIT CHARACTERISTICS ESTIMATION

Here also was provided to analyze the object state by all totality sources activities observation. This procedure is realized by analyzing unit BAU (Fig.4). For example, it can be based on the adaptive switchboard principle using [11], i.e., at the every analyzing step it is chosen the most active among totality sources (fig.5). Each sensor activity measure is taken as the  $i$ -th signal current deviation from its previous sampled value until analyzed moment. This deviation is normalized after its mean-square deviation value. Larger normalized deviation is equivalent to more active source. Each  $i$ -th difference value is prepared by subtraction of the sample stored in memory element  $ME_{2i}$  from its current sample from the sources outputs (points 2, 3). They are fixed by the  $i$ -th subtraction element  $SE_{2i}$  then passed through divider  $D_i$  to the activity estimation element AEE.

This element notes the  $k$ -th most active source as well as puts a control signal at its corresponding output. It allows rewriting the value of the most active source in analyzed moment in the  $k$ -th memory element  $ME_{2k}$  through an open switch  $SW_{3k}$ . This allowance is realized by opening a corresponding switch  $SW_{3k}$ . It is a preparation to the next activity analyzing procedure. The number of each source activities manifestations during analyzing interval  $T_\alpha$  is fixed in element AEE for the proper sampling program at this partly stationary interval formation. The sampling program of the current partly stationary object state is checked by estimation unit EU and passed to unit PSU (point 9).

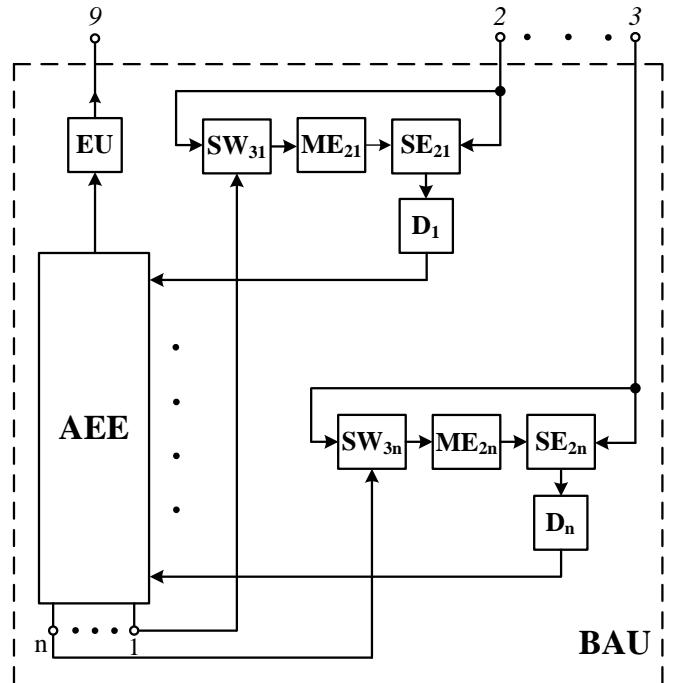


Fig.4. Behavior analyzing unit structure

In accordance with the principle of adaptive switchboard operation [11], the intensity of the  $i$ -th source  $\lambda_i$  is inverted to the average interval  $\bar{\tau}$  between two serial activity manifestations of the same  $i$ -th source. Let us consider that any current error  $\Delta(t)$  is described by random set of triangles (fig.6).

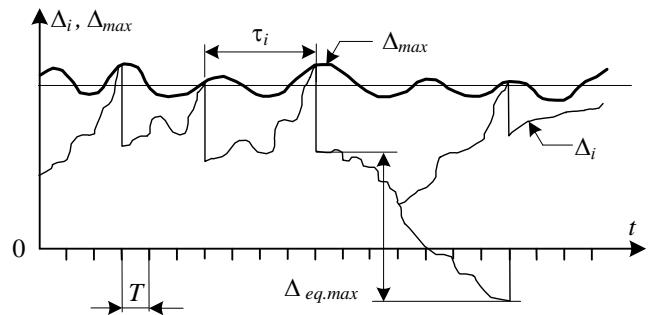


Fig.5. Formation of modulus of sampling errors maxima

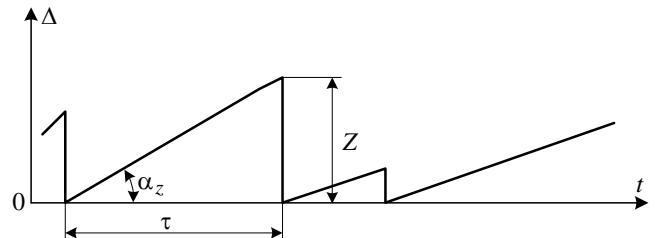


Fig.6. Dependence of sampling error in time

Interval  $\tau$  is defined as  $\tau = Z / tg\alpha_z$ . Since the random variables  $Z$  and  $tg\alpha_z$  are mutually independent, the approximate record looks like the following:

$$\bar{\tau} = \bar{z}/\bar{tg}\alpha_z \quad (4)$$

The value of the average tangent of angle  $\alpha_z$  is equal to an average value of source process derivative modulus as follow

$(\bar{tg}\alpha_z) = \left| \frac{\cdot}{\xi} \right|$ . There is known expression for a normal law of

the process distribution, i.e.,  $\left| \frac{\cdot}{\xi} \right| = \sigma_{\xi} \sqrt{\frac{2}{\pi}} = \sqrt{\frac{2}{\pi}} \omega_{li} \cdot \sigma_i$ , where

$\omega_{li}$  is a mean-square angular frequency (rad/s) of the  $i$ -th process;  $\sigma_i$  and  $\sigma_{\xi}$  are the standard deviation of the  $i$ -th process and its derivative, respectively. Namely,

$$\sigma_{\xi}^2 = \frac{1}{2\pi} \int_0^\infty \omega^2 G(\omega) d\omega, \quad \sigma_i^2 = \frac{1}{2\pi} \int_0^\infty G(\omega) d\omega. \quad (5)$$

The corresponding to the average interval  $\bar{\tau}$  between the two serial activity manifestations of the  $i$ -th source (4) at the adaptive sampling [11] intensity is described as follows:

$$\lambda_i = 1/\bar{\tau} = \sqrt{\frac{2}{\pi}} \cdot \frac{\omega_{li} \sigma_i}{C_2}, \quad (6)$$

here  $C_2$  is a constant value dependent on the frequency characteristics of measurement object sources totality and a synchronous channel tact.

For given equal probability of positive and negative current values of a sampling error, and therefore its equal zero expectation, one can write the expression for the mean square of the absolute value of the error, i.e.,

$$\overline{\Delta u_{si}^2} = \frac{1}{3} C_2^2. \quad (7)$$

The frequency of an adaptive switch analising procedure is defined by the sum of intensities from all sources of measurement object [11]. This value is used for analising unit BAU proper operation. Thus,

$$\frac{1}{T} = \sum_{i=1}^n \lambda_i. \quad (8)$$

If to take the regular type sampling interval of each source equal to the estimated by analizing unit BAU one, then after a comparison of expressions (6) and (15) it is stated that a sampling error is less at the adaptive serving than at the regular one (in  $\pi/2$  times).

#### IV. CONCLUSION

The regular serving procedure [12,13] is based on the number of everyone source activities obtained by BAU. As well as these results can be used for the entropy estimation of object state [14].

#### REFERENCES

- [1] B. Höfig; P. Eichinger; C. Richter, "Education 4.0 for Mechatronics – Agile and Smart", *Proceedings of the 18-th International Conference on Research and Education in Mechatronics (REM'2017)*, Wolfenbüttel, Germany, September 14 – 15, 2017, Wolfenbüttel: Published by

German Mechatronics Association, 2017 - DOI: 10.1109/REM.2017.8075250.

- [2] *Industry 4.0: the fourth industrial revolution – guide to Industrie 4.0* - <https://www.i-scoop.eu/industry-4-0/>
- [3] Georg E. Fanter, Paul Hegarty, Johannes H. Kindt, and Georg Schitter, "Data acquisition system for high speed atomic force microscopy", *Review of Scientific Instruments*, 76, 026118, 2005.
- [4] Johannes Gutleber, Steven Murray, Luciano Orsini, "Towards a homogeneous architecture for high-energy physics data acquisition systems", *Computer Physics Communications*, 153, pp.155-163, 2003.
- [5] S.Montebugnoli, G.Bianchi, L.Zoni, "Programmable fast data acquisition system", *Memorie della Societa Astronomica Italiana*, vol.10, pp.195-197, 2006.
- [6] Michael Rizk, Iyad Obeid, Stephen H. Callender, and Patrick D. Wolf, "A single-chip signal processing and telemetry engine for an implantable 96-channel neural data acquisition system", *Journal of Neural Engineering*, 4, pp.309-321, 2007.
- [7] Mark A. Davenport, Marco F. Duarte, Yonina C. Eldar and Gitta Kutyniok, "Introduction to Compressed Sensing," in *Compressed Sensing: Theory and Applications*, Y. Eldar and G. Kutyniok, eds., Cambridge University Press, 2011.
- [8] Rusyn B., Lutsyk O., Lukenyuk A., Pohoreliuk L., "Lossless Image Compresion in the Remote Sensing Applications", *Proceeding of the 2016 IEEE 1st International Conference on Data Stream Mining and Processing DSMP*, pp. 195-198, 2016.
- [9] O.Ivakhiv, A.Kowalczyk, R. Velgan, "Intelligent Programmable Measurement System", *Proceedings of the XVI IMEKO World Congress. Volume IX, topic 30 - Artificial Intelligence in Measurement Techniques*, Vienna-Wien, Austria, pp.341-345, September 25-28, 2000.
- [10] Orest Ivakhiv, Petro Mushenyk. Yuriy Hirnyak, "Intelligent Analyzing System", *Sensors & Transducers Journal*, Volume 24, Special Issue, P. 43-49, August 2013.
- [11] Kalahnikov I.D., Stepanov V.S., Churkin A.V. *Adaptive Data Gathering and Transmission Systems*. Moscow: Energiya Press, 1975, 240 p. (in Russian).
- [12] I.M. Teplyakov, I.D. Kalashnikov, and B.V. Roshchin, *Satellite Data Transmission Radio Links*. Moscow: Sovetskoe Radio Press, 1975 (in Russian).
- [13] O.Ivakhiv, A.Kowalczyk, R.Viblili, "Intelligent Measuring System Simulation", *Book of Abstracts. 16-th IMACS World Congress on Scientific Computation, Applied Mathematics and Simulation* (Lausanne - Switzerland, August 21-25, 2000), Ecole Polytechnique Federale de Lausanne, p.460.
- [14] Roman Velgan, Yuriy Hirniak, Orest Ivakhiv, Petro Mushenyk, Maksym Oleksiv. Entropy Estimation of Investigated Object State, *Proceedings of the XIII<sup>th</sup> International Conference Perspective Technologies and Methods in MEMS Design (MEMSTECH)*. Polyania, Ukraine, p.132-135, April 20-23, 2017.

# Algorithms of Landmark Robot Navigation Basing on Monocular Image Processing

Vasyl Koval

Department of Information Computing Systems and Control, Ternopil National Economic University, UKRAINE, Ternopil, 8 Chekhova str.,  
email: [vko@tneu.edu.ua](mailto:vko@tneu.edu.ua)

**Abstract:** The application of mobile robots is very important in environments that are dangerous or inappropriate for human life. One of the problems arising for the mobile robot when targeting point within the indoor application during navigation is the provision of its localization. In this paper, the developments of the algorithms that provide and enable mobile robot to position itself within the indoor environment by using one video camera and a landmark template is presented.

**Keywords:** mobile robot, robot navigation, navigation algorithm, mobile robot localization, landmark-navigation, indoor mobile robot navigation.

## I. INTRODUCTION

One of the most popular application for mobile robots (MR) is providing navigations in environments in which humans can't be present or environments that are dangerous to human's health [1,2]. The interaction of MR with the operating environment is provided by the application of a number of sensors for the perception of it, actuators (effectors) for influencing the environment and a control system that allows robot to perform purposeful and useful actions. By analyzing the indoor application of mobile robots, it is possible to conclude that its activities in the environment can be considered as a cyclic system.

Within the main loop, MR executes the procedures for the perception of the environment state, process the received information and determines actions that changes its position in the environment according to the fixed purpose. Thereafter, MR analyze changes and the information about the new environment state obtained is send to the control system. Due to these processes being executed, a new loop of mobile robot activities is organized till the purpose will be reached [3].

The actual problem is the creation of mobile robots that are capable of independently navigating and autonomously performing the assigned tasks. At the same time in most cases, humans provide remote control for the MR [4]. Such state is determined by the inability of the robot to make independent decisions and as a result, it provides number of shortcomings and increases the probability of erroneous actions. In addition, it is usually problematic for people to correctly assess the situation on a telemetry data basis and implementation of adequate control. These shortcomings can be avoided if the MR control by humans will be carried out at the level of goal setting, but not at the level of the task execution for individual movements. In this case, the robot must independently (or with minimal human impact) perform the assigned tasks. [5,6].

Typically, the technical vision system is used by MR during navigation. There are three strategic levels for reaching target point of movement by MR and they include: a) corresponding to the far, b) middle and d) near navigation. To be capable of providing such navigational levels, it is significant to develop some algorithms and tools that could support robot to estimate its position or localize at the operating environment.

## II. PROBLEM FORMULATION

One of the core task for robot's navigation is the determination of the MR position and orientation (often referred to as the pose) in its environment. The basic principles of landmark-based and map-based positioning also apply to the vision-based positioning or localization, which relies on optical sensors in contrast to ultrasound, dead-reckoning and inertial sensors.

Most localization techniques provide absolute or relative position and/or the orientation of sensors. Techniques vary substantially, depending on the sensors, their geometric models and the representation of the environment [7,8].

The geometric information about the environment can be given in the form of landmarks, object models and maps in two or three dimensions. A vision sensor should capture image features or regions that match the landmarks or maps.

The MR positioning means finding of the position and the orientation of a robot platform globally in the environment. Usually for this purpose of various types of range, finders are used. Finders have large numbers of drawbacks, the main one among them is that the finder can focus only on the configuration of the working area and the problem of localization (determining coordinates) is solved with errors.

Moreover, the traditional navigating systems usually use odometers for positioning of wheeled platform in an environment. They determine the path traversed by each of the wheels of robot. As a result, such approach leads to accumulated errors. Therefore, the practical problem is to create tools and algorithms that allow mobile robot to provide positioning for movement to the target.

Therefore, the ideal sensor for solving the distinct problems listed above is the video camera from the vision system of the robot. The proof of this statement may be human visual system. In this scientific report, the main attention focuses on the approaches that use photometric vision sensors, i.e., cameras for MR positioning.

In robotics, it is possible to find the implementation of stereo cameras for similar applications. Two or sometimes three cameras and special image processing techniques are used to reconstruct the robot's environment [9]. Stereo image

processing has its drawbacks. The main one among them is finding the correspondence between two stereo images that is very complicated. Moreover, the authors of these methods often simplify the process by creating artificial landmarks, including the usage of different kinds of structured lights, etc. At the same time, in nature there are many organisms that have successfully provided orientation in the environment by using only one video-sensor. This fact creates prerequisites for research methods for analog behavior in technical systems.

To address practical issues of the task definition, we will consider some of the environment in which mobile robot operates at industry (Fig. 1) [10].



Fig.1. The operating environment of mobile robot at industry [10]

In that environment, it is quite difficult to localize robot. Moreover, MR needs to determine its position independently for provision of the navigational goal, subsequently, to deliver the goods or perform other necessary operations.

Based on the above mentioned practical needs, it is proposed that the development of algorithms and software units use one camera and image processing technique to easily solve the task of positioning the mobile robot in the environment during the movement to the target.

### III. IDEA OF THE PROPOSED ALGORITHM

The idea of the Proposed Algorithm is to study and solve problems. Let us consider the technology that is taken from nature when organisms are oriented in space through various beacons (example: the sun). For this purpose, one of the possible solution of the previously mentioned task is proposed to fix a landmark on the ceiling of the technological environment where the mobile robot operates. In a situation wherein, the coordinates of landmark are known, there is a need for algorithms and software units of the image processing that will determine the robot's position in the environment.

Thus, the following geometric interpretation is proposed to solve this task (Fig. 2). According to Fig. 2, a special situation is proposed, when on the base of the mobile robot platform the video camera is fixed and directed vertically upward. Thus, the location of mobile robot platform determines by the position of video camera. The camera is located at some distance from the ceiling  $OM$  (Fig. 2). Any point located on the ceiling is projected through the center of the camera lens (point  $O$  on Fig. 2) to the sensor panel (plane  $ABDC$  on the Fig. 2). For localization of the mobile robot in the environment, the landmark template is fixed on the ceiling and has coordinates known as  $(X_2, Y_2)$ . To perform the movement of the mobile robot to its target, it is necessary to find its location in the environment (it is necessary to search the coordinates of points  $(X_1, Y_1)$  and the angle "Alfa") basing on the projections of

landmark to image. As part of the robot localization, one of the practical task is the identification of landmark on the image from the video camera.

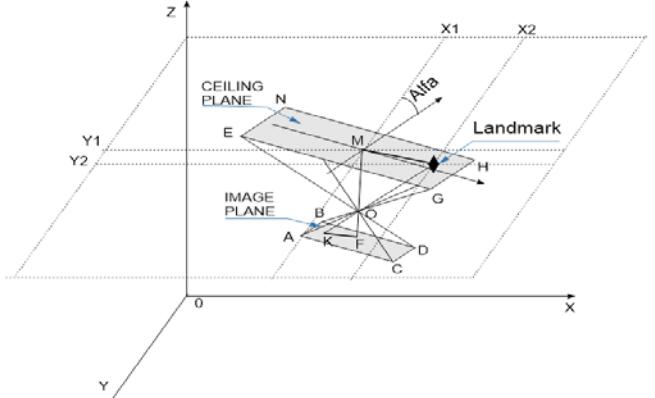


Fig.2. Geometric interpretation of the task definition

Thus, the input data for the developed algorithm and software units are color RGB images obtained by the robot's video camera. For the solution of the task that is given above, the following restrictions and assumptions are considered:

- a preliminary calibration of the camera was done. Due to the result of the video camera calibration, its position is fixed onboard of the mobile robot platform and does not change during operation;
- MR provides movement in a straight and flat horizontal surface like a floor, which practically represents a homogeneous coating (laminate flooring, linoleum, construction coupler);
- there are no overhanging objects that could cause a collision with a mobile robot in the environment;
- a landmark template exist in the environment with known parameters and it is visible for mobile robot;
- within presented restrictions, the operating three-dimensional model of the environment that is presented on Fig. 2 is considered.

The expected output of the algorithm is the selected segment of the landmark template at the image plane, which is used to calculate the trajectory of mobile robot movement to the target point.

Thus, to achieve this task, it is proposed to use the video camera as an effective passive sensor. By using this, the mobile robot will provide the proper positioning in the environment on the way to the target point.

### IV. GENERALIZED ALGORITHM OF MOBILE ROBOT NAVIGATION BASED ON MONOCULAR IMAGE

In general cases, the navigation of MR to the target is provided by using image processing from one camera. The robot navigation consists in analyzing of the current robot location and local targets that follow to global position. These local targets can be displayed as a line or landmarks representing a sequence of intermediate links that should follow the robot. Sometimes, there is a situation wherein the robot has only one global target to achieve. In this case, the movement of mobile robot must be ensured, taking into consideration the possible local obstacles or static architectural

designs of the environment. Local movement to the target may be provided by one of the known methods that are based on local or global navigation [11, 18]. Within this scientific work, the local movements of mobile robots are not considered, but it considers the algorithm by which the robot determines its position for predicting the direction of movement to the target as a subtask of robot navigation.

For simplicity in the consideration of the above presented principles of robot navigation, let consider the robot environment as grid-based model. In this environment, the coordinates of the target point are given, which represents the goal of the robot movement. The ultimate purpose of a robot navigation is to build a direction (trajectory) of movement to the global target point and to generate the control commands, which defines the required acceleration of MR wheels for maneuvering. The navigation task is completed when the robot is within a certain range with respect to the point of global goal.

For the provision of the above presented way of robot navigation, unlike existing known local methods of navigation [12-15], it is proposed to take the appropriate decisions for the direction of MR movement at each step in a loop. Thus, the decision to bypass obstacles and direction taken at each iteration of a loop depends on the location of the landmark on the image of the robot's camera. The main processes provided by mobile robot for navigation to its target can be presented by generalized algorithm and it consist of the following:

1. At the first step, there's an execution of the image processing procedures that initialize values for algorithms and provide camera calibration procedure.
2. At the second step of the algorithm, the position of the mobile robot platform (coordinates of its center point) and the target position are determined. The vector of the length between the point of the robot's position and the point of the target's position is determined (the distance to the target).
3. If the position of the mobile robot is within a certain radius delta (the concrete value is specified during the initialization procedure) from the target point, then it stops working and take a decision of reaching the goal of movement. In this case, the algorithm of mobile robot movement is finish. This moment represents the stop-point for the MR navigation algorithm.

Otherwise, the following sequence of steps are executed:

4. Gathering of the video-frame from the robot's video-camera.
5. It performs the segmentation of landmark template on the image received previously from the video-frame. Thereafter, the coordinates of the central point of the landmark template are calculated at the local coordinate system of the image.
6. It provides calculations of the directional angle of the Mobile Robot's position relative to the placement of landmark template that is segmented on the image from the video-camera.
7. It is performing the procedure of calculating the distance from the position of mobile robot to the central point of the landmark template.
8. It Performs the procedure of the MR positioning at the manipulating environment.

9. Based on the coordinates of the target and the position of the MR at the manipulating environment, it performs the procedure for defining the direction of movement.
10. Providing the MR maneuvering, based on the necessary parameters of acceleration for the MR motors. As a result, the movement of the robot's platform provides changes to its position in the environment (its coordinates).
11. Return to step two.

The flowchart of generalized algorithm is given on Fig. 3.

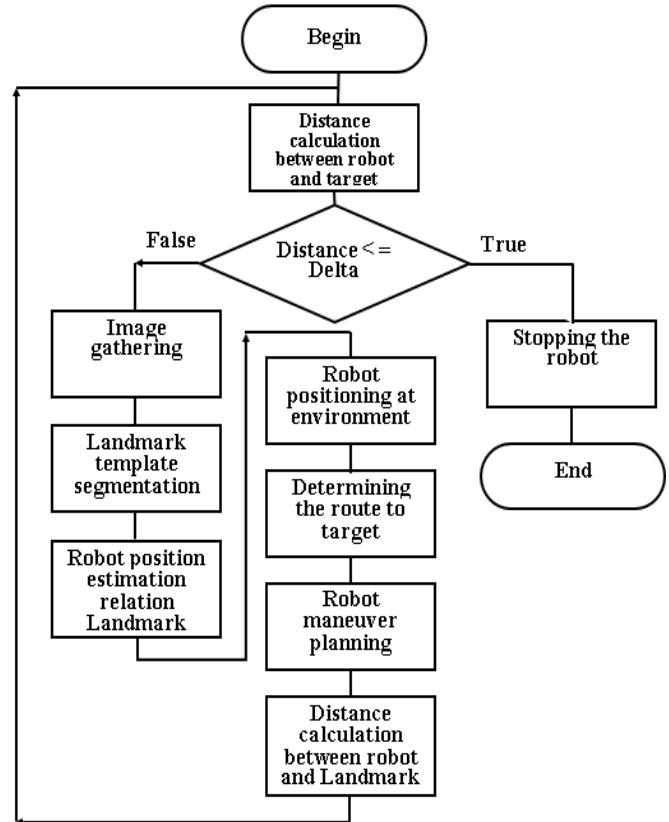


Fig.3. Generalized algorithm of mobile robot navigation

## V. LANDMARK TEMPLATE DETECTION ON THE IMAGE PLANE

During navigation, the MR estimate its location and position in the environment based on the landmark position. The last one, it is possible to receive based on image processing from one robot's video-camera. It means that the orientation of the given landmark of the images allows determination of the position of the robot's platform and as a result it provides smooth navigation.

In accordance to the list of steps presented above in the generalized algorithm, one of the first process of robot navigation is to capture a video frame from video camera. Performances of these steps can take place using existing and known possible approaches. At the same time, it is necessary to design the methods that can detect landmark template on video-image. To identify the landmark template for mobile robot navigation, it is proposed that algorithm is used to perform the following procedures (Fig. 4):

The RGB-image that receives from color video camera represents input for algorithm execution. The algorithm mentioned above selects some areas of pixels on the image (image segment) that belongs to possible landmark template that are selected. Thereafter, it is applying the procedure of rejecting all other than landmark template segments by using the various metrics. As the result of the algorithm, the image segment of pixels that respond to landmark template is selected.

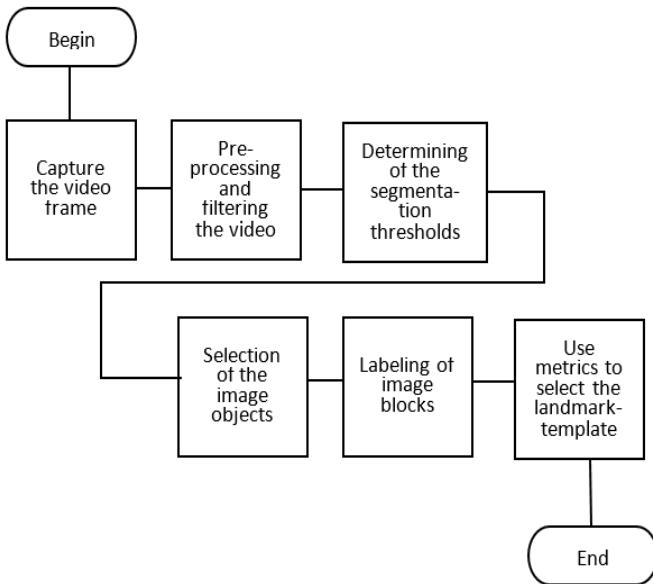


Fig.4. The algorithm for landmark template detection on the image plane

## VI. ALGORITHM IMPLEMENTATION

The above designed algorithms for robot navigation was explored by using Mat-lab software. The implementation of all processes that provide MR navigation to its target is currently under development. The conceptual interest of researches consists of obtaining the stable segmentation of landmark template for MR pose estimation in the environment. During researches, many navigation procedures could be implemented by the application of specific functions that are appropriate for the MR configurations individually, depending on the type of robot. For example, it could be possible to use ARIA environment for robots from ActivMedia Robotics Company [19, 20] (Fig. 5).



Fig.5. One video-camera application for navigation of mobile robot Pioneer P3-DX (potential application) [21]

Let consider the implementation of algorithm for landmark template detection on the image plane (Fig. 4) as the most important part for landmark robot navigation.

All the processes done were formalized mathematically for the investigation and implementation of the algorithms mentioned above. Also, it was designed specific graphical representation of the landmark template (Fig. 6a). The shape of landmark allows unique identification among other objects of the image, and to determine angular orientation on the global environment map. Additionally, the three metrics for guaranteeing the selection of landmark template among the other segments on the image plane were suggested:

- the number of pixels in the segment;
- the distance between the most remoted pixels in the segment;
- the presence and number of holes in the segment.

To investigate and demonstrate algorithm, a special situation of landmark location on the ceiling of MR environment was taken (Fig. 6a). As it could be seen on the image captured by the video-camera, there exist additional object (lighting lamp). Such object could be located at the range of the camera's vision and needs to be removed as unwished for processing. Median filter with 3x3 matrix operation were applied to each image pixel on Fig. 6a.

According to the algorithm of landmark detection, the following values of thresholds were selected for red, green and blue colors:  $R\_Tresh=75\pm28$ ,  $G\_Tresh=95\pm10$ ,  $B\_Tresh=133\pm10$ . The result of image thresholding presents on Fig.6b.

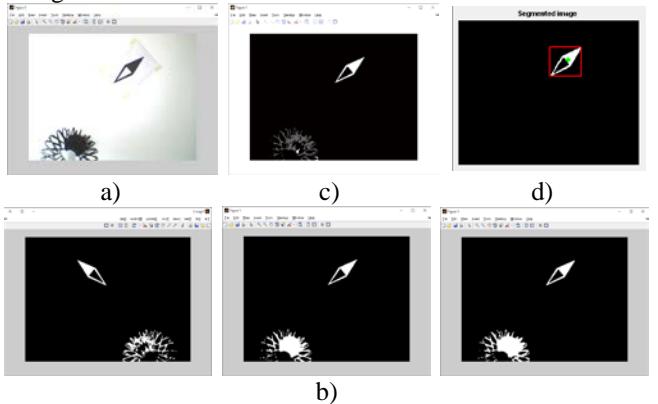


Fig.6. Results of image processing: a) the input image; b) the results of thresholding in the red, green and blue spectrums; c) result of labeling of segmented image (there are 58 objects indicated by different colors); d) the result after applying metrics

The initial image was binarized by combining of the segmented images on Fig.6b. During the processing of the neighbor-segmented pixels at the initial image, 58 segments were selected as candidates to be landmark template for MR navigation (Fig.6c).

Thereafter the following values were experimentally selected for proposed metrics: the number of pixels in the segment (2200-3120); the distance between the most remoted pixels in the segment (150±10); the presence and number of holes in the segment (3 holes).

The result of applying mentioned metrics in the proposed algorithm of detecting the landmark template on the image is shown on Fig.6d. As it was expected, only one segment among

the plurality of image objects were selected. It includes 2436 pixels and three holes. The distance between the most remote points is 147.

Experimental studies have shown the usage of a sufficient metric for identifying landmark template at 200 different locations at the environment.

The actual representation of the algorithm scenarios will be demonstrated during the presentation.

## VII. SUMMARY AND CONCLUSION

In this paper, algorithms of mobile robot movements were developed and experimentally investigated by using a one video camera. This practical task was reached by applying localization techniques using landmark template detection.

The generalized algorithm that allows mobile robot to move to the target was developed based on reading from one video camera and image processing procedures.

The graphic template of landmark was designed, which allows the MR to identify its position as the image among other objects and allows it to determine angular orientation on a global environment of mobile robot.

The algorithm for landmark template segmentation was designed based on the image processing that allows the MR to identify its position on the image plane. By knowledge of the position of landmark template in the environment, it is possible to localize mobile robot.

The experimental studies of the proposed algorithm of landmark template detection on the video images have shown the stability on each algorithm step and provided a selection of one segment among the plurality of image objects.

## REFERENCES

- [1] Robla-Gómez S., Becerra V., "Working Together. A Review on Safe Human-Robot Collaboration in Industrial Environments", *IEEE Access* (Vol. 5), November 14, 2017, pp. 26754–26773.
- [2] Baudoin Y., Habib M., "Using Robots in Hazardous Environments", 1st Edition, Woodhead Publishing, 2010, P. 692.
- [3] Evans, J., PatrUn, P., Smith, B., Lane, D.M., "Design and Evaluation of a Reactive and Deliberative Collision Avoidance and Escape Architecture for Autonomous Robots", *Autonomous Robot* Vol. 24, 2008, pp. 247–266.
- [4] Goebel S, Jubeh R, Raesch S-L & Zuendorf. A. "Using the Android Platform to control Robots", In *Proceedings of 2nd International Conference on Robotics in Education (RiE 2011)*. Vienna, Austria, September 2011, pp. 135–142.
- [5] Siegwart, Roland, Nourbakhsh, Illah Reza, "Introduction to Autonomous Mobile Robots (Intelligent Robotics and Autonomous Agents series)" / Siegwart, Roland, Nourbakhsh, Illah Reza, Scaramuzza, MIT Press; 2nd Revised edition, 2011, P.453
- [6] Siciliano Bruno, Khatib Oussama "Springer handbook of robotics", Springer International Publishing, 2016, P. 2227.
- [7] Arras, K.O., Castellanos, J.A., Schilt, M., Siegwart, R., "Feature-based Multi-hypothesis Localization and Tracking Using Geometric Constraints", *Robotics and Autonomous Systems* 44, 2003, pp. 21-53.
- [8] Betke Margrit, Gurvits Leonid "Mobile robot localization using landmarks", *IEEE Transaction on robotics and automation*, Vol 13, No. 2, April 1997, pp.251-263.
- [9] H. Roth, A. Sachenko, V. Koval, O. Adamiv, V. Kapura, "Evaluation of Camera Calibration Methods for Computer Vision System of Autonomous Mobile Robot", *Proceedings of International Conference "Modern Information and Electronic Technologies" (MIET-2009)*, Odessa (Ukraine), 2009, p. 29.
- [10] "Are Robots About to Take Over E-commerce Warehouses?", 2018, <http://www.airindknows.com/are-robots-about-to-take-over-e-commerce-warehouses/>.
- [11] Jian, Y., "Comparison of Optimal Solutions to Real time Path Planning for a Mobile Vehicle ", by Y. Jian, Q. Zhihua, W. Jing, C. Kevin, *IEEE Transactions on Systems, Man and Cybernetics*, Part A, System and Humans, Vol. 40, 2010, pp. 721–725.
- [12] Ersson T., Hu X., "Path Planning and Navigation of Mobile Robots in Unknown Environments", *IEEE Journ. of Robotics and Automation*, # 6, 2010, pp. 212–228.
- [13] J.L. Guzmán, M. Berenguel, F. Rodríguez, and S. Dormido, "MRIT: Mobile Robotics Interactive Tool" [electronic resource], 2018, <http://aer.ual.es/mrit/>.
- [14] O. Adamiv, V. Koval, V. Dorosh, G. Sapozhnyk, V. Kapura, "Mobile Robot Navigation Method for Environment with Dynamical Obstacles", *Proceedings of the IEEE Fifth International Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications*, 21-23 September 2009, Rende (Cosenza), Italy, pp. 515-518.
- [15] Chernonozhyn, V.A., "Local Area Navigating System for ground mobile robots", *Scientific and Technical Journal YTMO St. Petersburg State University*, 2008, №57, pp. 13-22.
- [16] Oleh Adamiv, Vasyl Koval, Arunas Lipnickas, Viktor Kapura, "Local navigation method for improvement of mobile robot movement", *Proceedings of the 3rd International Conference Mechatronic Systems and Materials (MSM 2007)*, Kaunas (Lithuania), 2007, pp. 245- 246.
- [17] William Benn and Stanislao Lauria, "Robot Navigation Control Based on Monocular Images: An Image Processing Algorithm for Obstacle Avoidance Decisions", *Hindawi Publishing Corporation Mathematical Problems in Engineering*, Volume 2012, P.14.
- [18] Olivier Koch, Matthew R. Walter. "Ground Robot Navigation using Uncalibrated Cameras", In *Proc. IEEE International Conference on Robotics and Automation (ICRA)*, May 2010, pp. 2423-2430.
- [19] Adept Mobile robots, 2014; <http://www.activmedia.com/>.
- [20] "AmigoBot Operations Manual, revision 4.3", 2018, <http://robots.mobilerobots.com/wiki/Manuals>.
- [21] "Mobile robotics platforms", 2018, [https://raweb.inria.fr/rapportsactivite/RA2015/lagadic/ui\\_d51.html](https://raweb.inria.fr/rapportsactivite/RA2015/lagadic/ui_d51.html).

# Development and Research of Conveyor Structures of Binary Number Sorting Algorithms

Volodymyr Gryga<sup>1</sup>, Yaroslav Nykolaichuk<sup>2</sup>, Natalia Vozna<sup>2</sup>, Artur Voronych<sup>3</sup>, Boris Krulikovskyi<sup>4</sup>

1. Department of Computer Engineering and Electronics, Vasyl Stefanyk Precarpathian National University, UKRAINE, Ivano-Frankivsk, 57 Shevchenko str., email: [v.dr\\_2000@ukr.net](mailto:v.dr_2000@ukr.net)
2. Department of Specialized Computer Systems, Ternopil National Economic University, UKRAINE, Ternopil, 8 Chekhova str., email: [ya.nykolaichuk@tneu.edu.ua](mailto:ya.nykolaichuk@tneu.edu.ua), [n.vozna@ukr.net](mailto:n.vozna@ukr.net)
3. Department of Computer Systems and Networks, Ivano-Frankivsk National Technical University of Oil and Gas, UKRAINE, Ivano-Frankivsk, 15 Karpatksa str., email: [a.voronych@it.nung.edu.ua](mailto:a.voronych@it.nung.edu.ua)
4. Department of Computer Engineering, National University of Water Management and Nature Resources Use, UKRAINE, Rivne, 11 Soborna str., email: [kboris@ukr.net](mailto:kboris@ukr.net)

**Abstract:** The characteristics of the structures complexity of conveyor sorting devices, constructed on the basis of parallel algorithms for sorting binary arrays are analyzed. The improved structure of the conveyor sorting device, which compared to the known one had 1.9 times less hardware complexity and 1.5 times higher performance, was developed and researched. Modeling of VHDL models of conveyor sorting devices was carried out, their synthesis and implementation was made in the Xilinx FPGA by automated designing Vivado system applications. The received practical results of the complexity characteristic of the developed conveyor devices coincide with theoretical calculations. There are also determined the relevant areas of application of the developed devices.

**Keywords:** special device for sorting arrays of data, conveyor sorting device, basic sorting elements, comparison scheme.

## I. INTRODUCTION

Sorting is one of the common problems of data processing and it is generally understood as a problem of placing elements of disordered set of data sets values in order of monotonic increase or decrease [1]. The sorting operation takes on average 25% of machine time [2] and it is most often used in the tasks of digital processing of signals and images, similar to convolution operations, fast Fourier transform, and others [3]. We know many methods of sequential and parallel sorting of binary numbers [1-9]. There is a large number of algorithms to organize sorting in modern digital signal processors, and each of these algorithms has its advantages and disadvantages. Performing a software sorting operation is time-consuming and generally used to perform consistent data sorting methods. Particularly effective is the parallel performance of algorithm sorting operations using hardware methods that significantly accelerate the execution time of the algorithm. Such methods of parallel sorting algorithms, in which the sequence of executed operations only depends on the number of input data, are the following: the Batcher's method [2], the modified "bubbles" method [1],

the "pairwise-odd" permutation method, the fusion method and other [3].

If we use the hardware method, then there is ensured implementation of the binary number sorting operation in real time and the possibility of applying new circuit design solutions with the use of a modern element base with an orientation towards implementation in the FPGA form.

The implementation of highly efficient conveyor devices for sorting binary arrays requires extensive use of a modern element base, the improvement of existing and the development of new methods, algorithms and device structures.

In this regard, an important task is to develop efficient structures for conveyor sorting devices (CSDs), which will improve the time characteristics of multithreaded data processing. The implementation of such devices in the languages of the hardware description and their synthesis in FPGA will enable the designer to choose the optimal hardware cost and time-efficient characteristics of the structure of the CSD.

## II. CONVEYOR STRUCTURES ANALYSIS OF PARALLEL ALGORITHMS FOR BINARY NUMBERS SORTING

The conveyor processing principle involves the alignment in time operators of the algorithms sorting on different data [5-8].

Conveyor structures of sorting devices of binary arrays is developed and analyzed below. They are based on the known parallel sorting algorithms presented in a graphical form, and their hardware and time complexity are estimated. (Their system characteristics)

The hardware implementation of the known parallel sorting algorithms in the conveyor structures involves the full reflection of their flow graphs algorithm (tiered-parallel algorithm forms) [4-8,10,11,13] into the structure of the operating device. The vertices of the graph (functional operators) will correspond to the hardware block (operation) and the arcs will correspond to lines for the transmission of input data of intermediate and final results. The conveyor registers are placed on the lines that connect the operating

units of one tier with the operating units of the previous tier.

Tier-parallel algorithm forms determines the degree of parallelism of the graph (the maximum number of vertices in a single tier or the width of the graph), as well as the minimum-possible time of calculation of the given algorithm (number of tiers or height of the graph) [6-8].

It is shown the structure of the CSD in Fig. 1, which is based on the numbers sorting by modified "bubbles" sorting method [1,13].

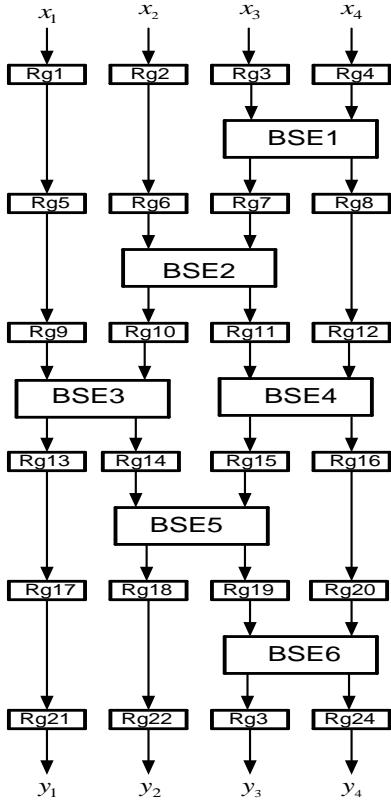


Fig.1. The structure of Conveyor sorting device for 4 values by the modified "bubbles" sorting method.

The basis for the above-presented structure of conveyor sorting device is the basic sorting elements (BSE) and conveyor register (Rg).

The structure of this conveyor device consists of the same basic sorting operations (max/min). CSD structure (Fig. 1) contains  $N(N-1)/2$  basic sorting operations and  $N(2N-3)$  conveyor registers for input  $N$  values.

The time complexity of this conveyor device structure is determined by the critical distribution of the signal through  $(2N-3)$  basic operations of sorting and  $(2N-3)+1$  conveyor registers.

It is shown the structure of the CSD in Fig. 2, which is constructed on the basis of sorting algorithm by "pairwise-odd" permutation method [2,13].

CSD structure (Fig. 2) contains  $N(N-1)/2$  basic sorting operations and  $(N \times N)$  conveyor registers for  $N$  input values.

The time complexity of this conveyor device structure is determined by the critical distribution of the signal through  $N$  basic operations of sorting and  $(N+1)$  conveyor registers.

It is shown the CSD structure in Fig. 3, which is constructed on the basis of the sorting algorithm graph by Batcher's method [3,13].

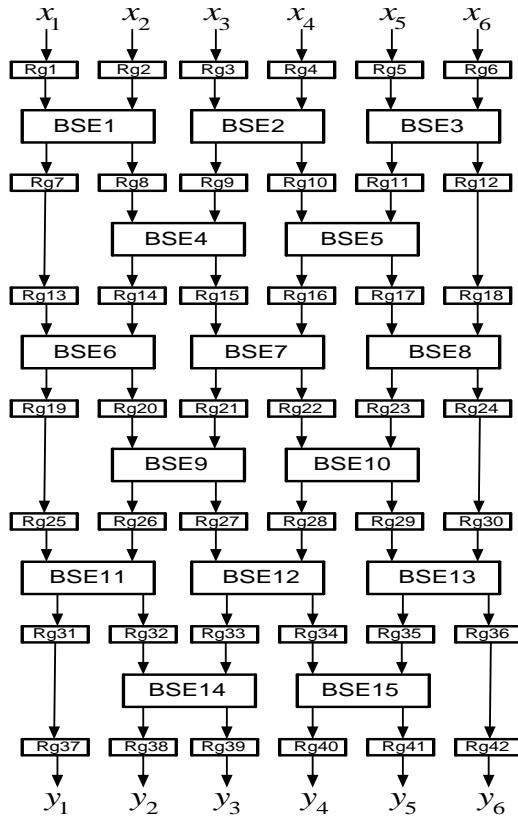


Fig.2. The structure of Conveyor sorting device for 6 values by "pairwise-odd" permutation method.

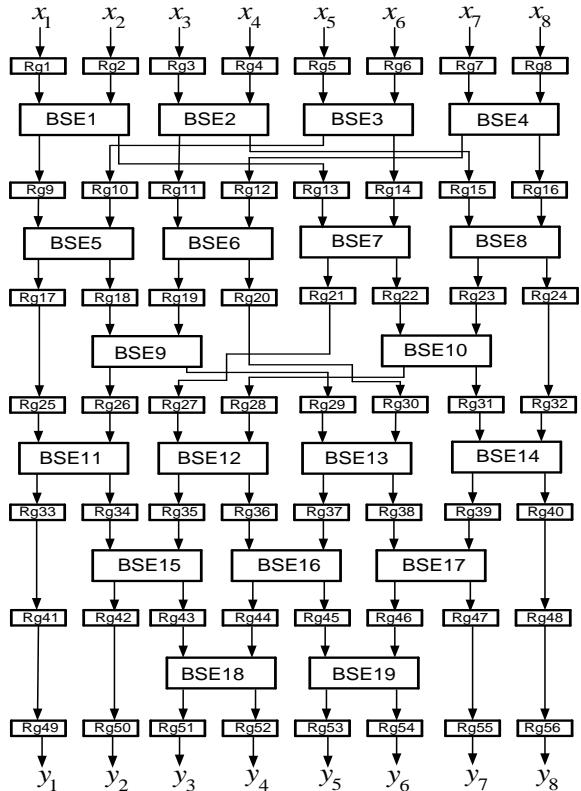


Fig.3. The structure of Conveyor sorting device for 8 values by Batcher's method.

CSD structure (Fig. 3) contains  $0,48N\ln^2 N$  basic sorting

operations and  $\frac{1}{2}(\log_2 N)([\log_2 N + 1]) \times N$  conveyor registers for  $N$  input values.

The time complexity of this conveyor device structure is determined by the critical distribution of the signal through  $\frac{1}{2}(\log_2 N)(\log_2 N + 1)$  basic operations of sorting and  $(N-1)$  conveyor registers.

The conveyor structures of the sorting devices (Fig. 1-3) consist of the same type of basic sorting elements that compare two numbers and more of them are sent to the first output and less to the second output. The internal structure of sorting basic element is presented in Fig. 4.

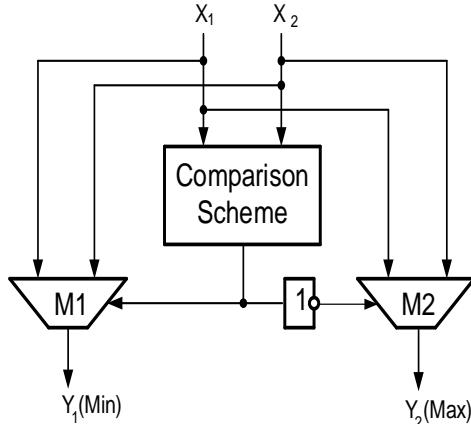


Fig.4. Internal structure of the basic sorting element

When comparing two numbers ( $X_1, X_2$ ) in the comparison scheme "for more" ( $X_1 > X_2$ ), the output of the scheme is formed a comparison sign, the direct value of which controls the multiplexer M1, to issue a smaller number to the output ( $Y_1$ ) and the inverse value controls the multiplexer M2 to issue a larger number at the output ( $Y_2$ ) [5,13].

The internal structure of the comparison scheme can be constructed on the basis of logical elements [5] or accelerated carry adders [13].

The calculation of the hardware and time complexity of different comparison schemes and 2-input multiplexers is given in paper [6].

With the account to these data, the hardware complexity of the CSD by the modified "bubbles" method for  $N = 8$  and  $n = 4$  equals:

$$A_{CCSD} = N \times (N-1)/2 \times (A_{CS} + 2 \times A_{MP}) + (n \times A_{REG} \times L_{REG}) = 28 \times (24 + 56) + (4 \times 4 \times 112) = 4032 \text{ (gates)},$$

where  $A_{CCSD}$  - hardware complexity of classical CSD,

$A_{CS}$  - hardware complexity of comparison scheme,

$A_{MP}$  - hardware complexity of multiplexer,

$A_{REG}$  - hardware complexity of register,

$L_{REG}$  - quantity of conveyor register.

The time complexity of this conveyor device equals:

$$\tau_{CCSD} = (\tau_{CS} + \tau_{MP}) + (14 \times \tau_{REG}) =$$

$$= (6 + 3) + (14 \times 2) = 37v \text{ (micro-cycles)},$$

where  $\tau_{CCSD}$  - time complexity of classical CSD,

$\tau_{CS}$  - time complexity of comparison scheme,

$\tau_{MP}$  - time complexity of multiplexer,

$\tau_{REG}$  - time complexity of register.

### III. DEVELOPMENT OF AN IMPROVED STRUCTURE OF THE CONVEYOR SORTING DEVICE

The bubble CSD (Fig. 1) can be improved by performing independent base-sorting operations separately for the first and second half of the input data on the two conveyor structures.

Values  $X_i$  are already ordered by recession and need to be further applied to  $((N/2)^2 - N/2) / 2 + N/2$  basic sorting elements and  $4N$  conveyor registers.

It is shown an improved CSD for 4-bit binary numbers ( $n = 4$ ) for 8 input values ( $N = 8$ ) in Fig. 5.

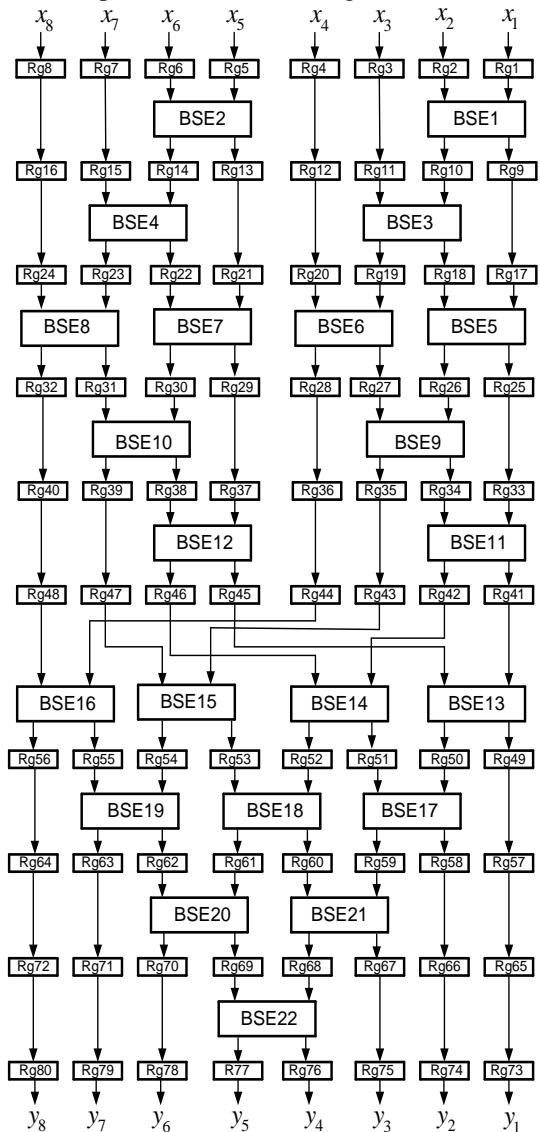


Fig.5. An improved structure of the CSD

The number of basic sorting elements for this CSD for the binary numbers is equal to  $3((N/2)^2 - N/2) / 2 + N/2$ , and

the number of conveyor registers is  $6N + 4N$  for  $N$  input values.

The hardware complexity of an improved CSD for  $N = 8$  and  $n = 4$  is equal to:

$$\begin{aligned} A_{ICSD} &= 22 \times (A_{CS} + 2 \times A_{MP}) + (n \times A_{REG} \times L_{REG}) = \\ &= 22 \times (16 + 24) + (4 \times 4 \times 80) = 2160 \text{ (gates)}, \end{aligned}$$

where  $A_{ICSD}$  - hardware complexity of improved CSD,

$A_{CS}$  - hardware complexity of improved comparison scheme [13],

$A_{MP}$  - hardware complexity of improved multiplexer [13],

$A_{REG}$  - hardware complexity of register,

$L_{REG}$  - quantity of conveyor register.

The time complexity of such CSD is equal to:

$$\begin{aligned} \tau_{ICSD} &= (\tau_{CS} + \tau_{MP}) + (10 \times \tau_{REG}) = \\ &= (3 + 2) + (10 \times 2) = 25 \nu \text{ (micro-cycles)}, \end{aligned}$$

where  $\tau_{CCSD}$  - time complexity of classical CSD;

$\tau_{CS}$  - time complexity of comparison scheme;

$\tau_{MP}$  - time complexity of multiplexer;

$\tau_{REG}$  - time complexity of register.

So, in comparison with the classic device, we get a reduction in the hardware complexity in  $K_A = 4032/2160 = 1,9$  times and increase the speed in  $K_\tau = 37/25 = 1,5$  time.

#### IV. RESULTS OF THE CONVEYOR SORTING DEVICES RESEARCH

It is shown a dependence graph of logical elements (valves) number on the input data number for the conveyor structures of the known (classical) and improved sorting devices by the modified "bubbles" method.

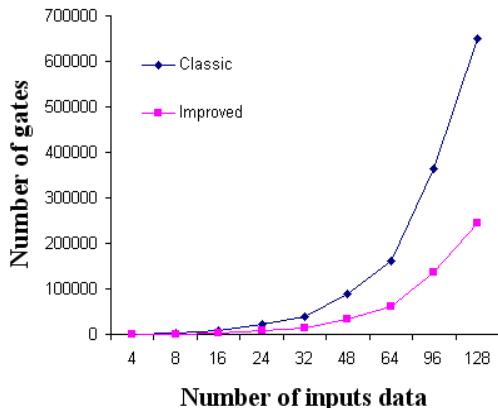


Fig.6. Graph of the dependence of the number of valves on the input data number for the CSDs

The graph shows that the number of logical elements for an improved CSD is 1.9 times less than for a classic CSD.

It is shown a graph of the dependence of the time complexity in Fig. 7 expressed in the microtacts on the input data value for the structures of the known (classical) CSD and the improved CSD with the modified "bubbles" method.

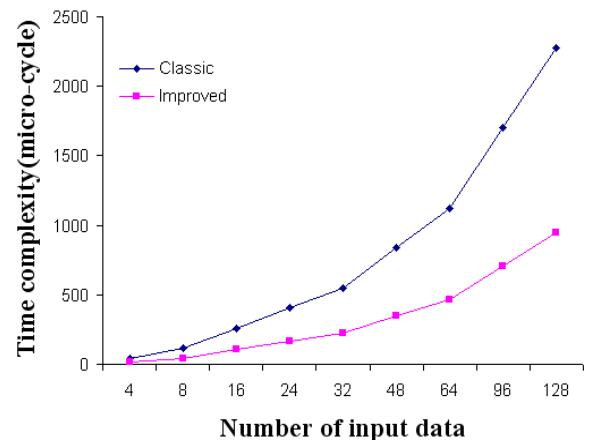


Fig.7. The dependence graph of the time complexity on the number of the input data for the CSDs

It is shown in the graph that the time complexity of an improved CSD requires about 1.5 times less microtacts than the classical CSD.

#### V. RESULTS OF CSD's SIMULATION AND SYNTHESIS ON FPGA

Structures of classical and improved CSD for 8 input one-byte numbers are described in VHDL (Virtual Hardware Description Language). The simulation of the developed CSDs on the functional level was carried out, their RTL circuits were obtained and the Xilinx FPGA synthesis was performed.

It is shown in Fig. 8 a functional diagram of the improved CSD operation for 8 input one-byte numbers.

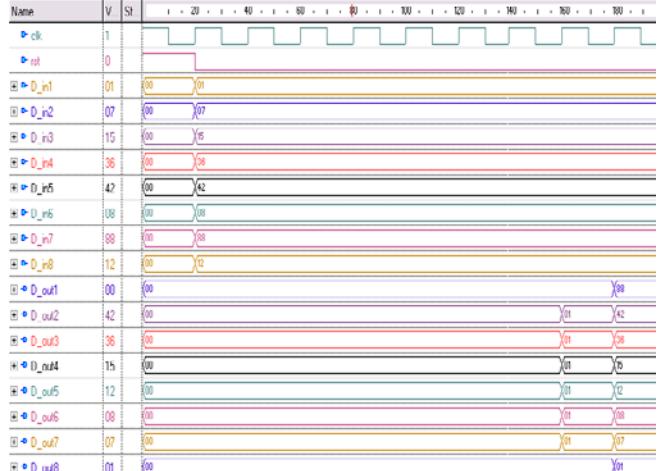


Fig.8. Functional simulation diagram for CSD of 8 one-byte numbers

The diagram shows that an array of unsorted 8-bit values is given at the inputs of the CSD ( $D_{in1}, \dots, D_{in8}$ ) on the 1<sup>st</sup> cycle. Then we get a sorted descending order at the outputs ( $D_{out1}, \dots, D_{out8}$ ) at the 9<sup>th</sup> cycle.

It is shown in Fig. 9 the topology view of improved CSD with the VHDL implementation model on the xc7a100tcsg324-1 crystal (Artix-7 family) of the Xilinx firmware by Vivado CAD.

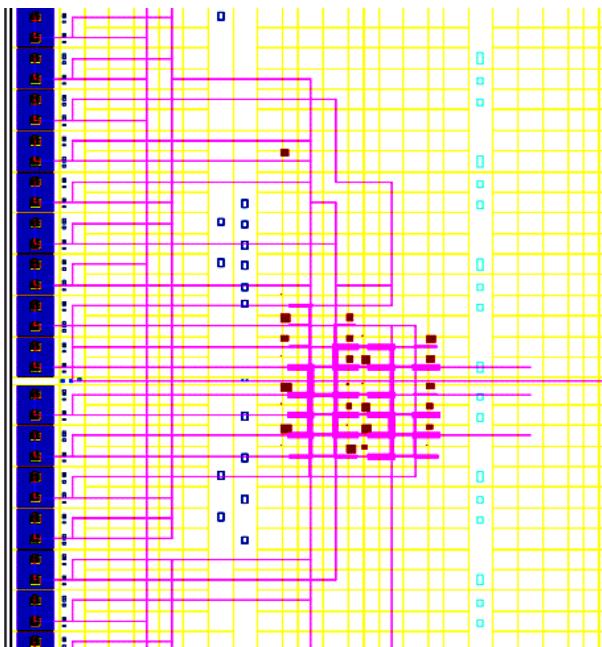


Fig.9. View of the crystal topology of CSD during implementation at the FPGA

Vivado CAD tools placed the implemented VHDL project of the improved CSD almost in the center of the crystal.

Table 1 presents the results of the synthesis of implemented CSDs for sorting 8 one-byte numbers on the FPGA of the Xilinx firmware.

TABLE 1. RESULTS OF THE CSDS SYNTHESIS ON FPGA

| FPGA      | Classic CSD                   |                          | Improved CSD                  |                          |
|-----------|-------------------------------|--------------------------|-------------------------------|--------------------------|
|           | Blocks quantity<br>FPGA (CLB) | Clock frequency<br>(MHz) | Blocks quantity<br>FPGA (CLB) | Clock frequency<br>(MHz) |
| 1 Artix 7 | 950                           | 109,9                    | 475                           | 174,7                    |

As can be seen from Table 1, experimental results coincide with analytical calculations.

## VI. CONCLUSION

During the research of CSDs it is determined that they are widely used in multithreaded data processing, and are often used in digital processing of signals, images and sorting networks for the rapid transfer of large data arrays.

The improved structure of the CSD of binary arrays by the modified "bubble" method was developed, the hardware and time complexity calculation was performed.

As a result of the comparison with the classical CSD for binary number array by the modified bubble method it is received decreasing the cost of the equipment in 1.9 times and increasing of speed in 1.5 times, which is confirmed by the results of the practical implementation of the Xilinx FPGA.

## REFERENCES

- [1] D. Knuth, The art of computer programming, V.3: *Transl. from eng.* M.: "Peace", pp. 841, 1978.
- [2] Thomas Braunl., Parallel Programming: *Transl. from german*, K.: Vyscha Shkola., pp. 358, 1997.
- [3] S.Y. Kung, VLSI array processors. *Trans. from English - M.*: "Peace", pp. 672, 1991.
- [4] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest and Clifford Stein. Introduction to algorithms, 3rd edition, Massachusetts Institute of Technology, pp.1313, 2009.
- [5] A. Melnyk, Memory with ordered access: monograph, Lviv: Press Lviv Polytechnic, pp. 296, 2014.
- [6] A. Melnyk, Computer Architecture, Lutsk:Regional publisher, pp. 470, 2008.
- [7] V. Voevodin, Parallel Computing. Textbook for high schools, BHV:Peter, pp. 608, 2002.
- [8] Andrew S. Tanenbaum Computer Architecture Austin. 6<sup>th</sup> ed., St. Petersburg: "Peter", pp. 816, 2013.
- [9] O. Kehret, A. Walz and A. Sikora, "Integration of hardware security modules into a deeply embedded TLS stack" *International Journal of Computing*, vol. 15, issue 1, pp. 24-32, 2016.
- [10] V. Zadiraka, Ya. Nykolaichuk. Methods of effective protection of information flows, Ternopil: *Terno-graf*, pp.308, 2014.
- [11] S. Melnychuk, A. Voronych, L. Nykolaichuk and B. Krulikovskyi "The Structure and Components of Embedded Special Processors for Determination of Entropy Signals and Random Massages" *Perspective technologies and methods in MEMS design. Proceedings of XIIIth International Conference. MEMSTECH 2017*, Lviv-Svalyava, Ukraine, pp.81-84, February, 2017.
- [12] B. Krulikovskyi, A. Davletova, V. Gryga and Y. Nykolaichuk "Synthesis of Components of High Performance Special Processors of Execution of Arithmetic and Logical Operations Dsts Processing in Theoretical and Numerical Basis Rademacher", *Proceedings of XIVth International Conference. CADSM'2017*, Lviv-Poljana, Ukraine, pp. 214-217, February, 2017.
- [13] V. Gryga, Y. Nykolaichuk, N. Vozna and B. Krulikovskyi "Synthesis of a microelectronic structure of a specialized processor for sorting an array of binary numbers" *Perspective technologies and methods in MEMS design. Proceedings of XIIIth International Conference. MEMSTECH 2017*, Lviv-Svalyava, Ukraine, pp.170-173, April, 2017.
- [14] V. Gryga, I. Kolosov and O. Danyluk "The development of a fast iterative algorithm structure of cosine transform" The Experience of Designing and Application of CAD System in Microelectronics. *Proceedings of XIIIth International Conference. TCSET'2016*, Lviv-Poljana, Ukraine, pp.506-509, February, 2016.
- [15] R. Dunets and V. Gryga "Spatio-temporal synthesis of transformation matrix of reverse fast cosine transformation", *Proceedings of XIIIth International Conference. CADSM'2015*, Lviv-Poljana, Ukraine, pp.45-49, February, 2015.

# Synchronous Rectification in High-Frequency MagAmp Power Converters

Volodymyr Yaskiv<sup>1</sup>, Anna Yaskiv<sup>1</sup>, Oleg Yurchenko<sup>2</sup>

1. Department of Applied Information Technologies and Electrical Engineering, Ternopil Ivan Puluj National Technical University, UKRAINE, Ternopil, 56 Ruska str., email: [yaskiv@yahoo.com](mailto:yaskiv@yahoo.com)
2. The Institute of Electrodynamics of The National Academy of Sciences of Ukraine, UKRAINE, Kyiv, 56 Peremogy ave., email: [yuon@ied.org.ua](mailto:yuon@ied.org.ua)

**Abstract:** The paper describes new approaches to high-efficient high-frequency power supply design for specialized computer systems, which require high load current at low output voltage. It is suggested to use semiconductor power converters based on high-frequency magnetic amplifiers. Paper shows the ways to increase converter's efficiency due to the use of a synchronous rectifier based on MOSFETs.

**Keywords:** high-frequency magnetic amplifier, synchronous rectifier, power converter, high level of the load current.

## I. INTRODUCTION

Modern specialized computer systems require high-quality and high-efficient power supply for their proper functioning. Reliability of such system is determined first of all by the reliability of its power supply. One of the peculiarities of specialized computer systems is that they consume high currents (dozens, often hundreds amperes) at low input voltage (3.3V, 5V, etc.) This results in strict requirements to power converters that are used as power supplies for such systems. They include high level of reliability, efficiency, specific power along with high quality of output voltage and its dynamic characteristics in the whole range of change of the load current. Moreover, operation of such power converters should cause the lowest possible level of both conductive and radiative electromagnetic interferences.

Nowadays, power supplies for specialized computer systems are realized as high-frequency power converters. Their efficiency is mostly defined with the operation modes of the high-frequency output rectifier, as the largest part of power converter losses at high level of load current is caused by the rectifier diodes.

Technical characteristics of modern rectifying diodes (including Schottky diodes) for low output voltage applications allow to provide satisfactory efficiency for high-frequency semiconductor low and medium power DC power converters.

With the appearance of high-frequency MOSFETs develops a new rectifier topology: synchronous rectifier. Its novelty consists in the use of a MOSFET instead of a rectifying diode, which is controlled in a function of voltage of high-frequency power transformer secondary winding (synchronously with this voltage) [1-3]. Works [4-7] describe digital solutions for synchronous rectifier control. The current paper introduces a simplified topology of a power converter

with synchronous rectifier which does not require digital controllers.

Modern semiconductor component manufacturers specify MOSFETs for synchronous rectifiers as a separate category, and work on decreasing their channel resistance in the conducting state. For instance, this parameter equals 0.2 mΩ for recent synchronous rectifier MOSFETs by International Rectifier [8].

## II. FUNDAMENTALS OF MAGAMP POWER CONVERTERS DESIGN

MagAmp is just a coil wound on a core of amorphous alloy with a relatively rectangular hysteresis loop (fig. 1) [9-11]. A MagAmp, used as a switch, can block and delay the applied voltage. However, MagAmp cannot interrupt the current once started. Hence, MagAmps are used in pulse circuits where they are assisted by diode rectifiers, which cut off the current as the applied voltage changes polarity.

When the voltage of negative polarity is applied to MagAmp, its core demagnetizes (corresponds to 1-2 slope in fig. 1; t<sub>1</sub>...t<sub>2</sub> in fig. 2). The MagAmp core is unsaturated and due to high resistance there flows no current through its winding. When the input voltage changes its polarity to positive, MagAmp requires a certain volt-sec, which is the integral of voltage over time, to be applied to its terminals for the magnetic flux to build up in the core and reach the saturation level (interval 2-3 in fig. 1; t<sub>2</sub>...t<sub>3</sub> in fig. 2).

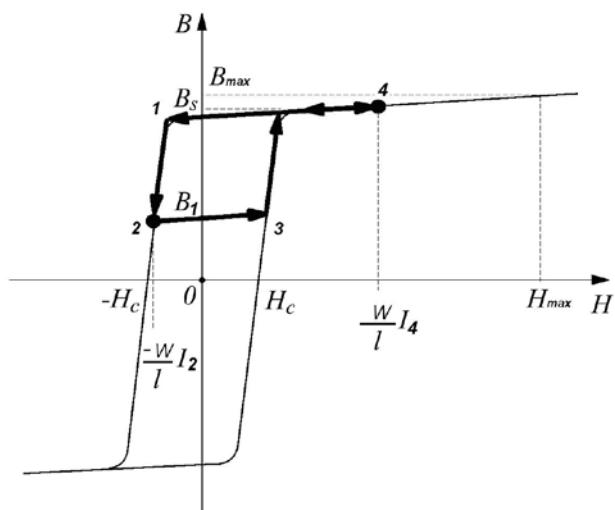


Fig. 1. Hysteresis loop of MagAmp core material

When the magnetic inductance reaches the saturation level (slope 3-4 in fig. 1), the MagAmp resistance approaches zero, which allows the current to flow through MagAmp's winding (interval 4-1 in fig. 1;  $t_3 \dots t_4$  in fig. 2).

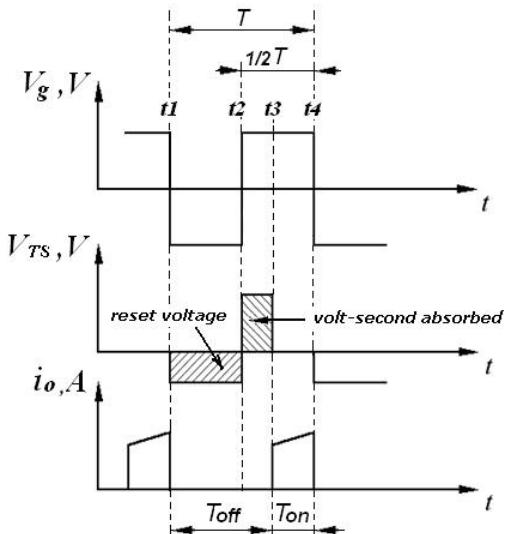


Fig. 2. Ideal MagAmp waveforms:  $V_g$  – transformer secondary winding voltage,  $V_{MS}$  – MagAmp switch voltage;  $i_o$  – MagAmp switch output current

In fig. 3 there is presented a functional scheme of a DC voltage regulator based on high-frequency magnetic amplifiers, which contains an unregulated high-frequency transistor voltage inverter 1, power transformer, push-pull centre-tapped rectifier, controlled MagAmps, output filter 2, load, control circuit 3, demagnetizing diodes [11].

However, providing a high level of output current in such voltage regulator is followed by the increase of losses on the diodes of high-frequency rectifier. This leads to a significant decrease of efficiency.

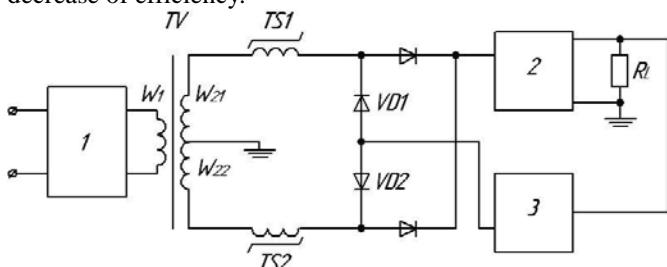


Fig. 3. Functional scheme of a DC voltage regulator based on high-frequency magnetic amplifiers

### III. MAGAMP POWER CONVERTER WITH SYNCHRONOUS RECTIFIER

It is suggested to substitute diodes of the output rectifier with MOSFETs, with a possibility of their synchronous control from the respective high-frequency transformer secondary windings. The functional scheme of such power converter is shown in fig. 4 [12].

The DC voltage regulator operates in the following way. When the control half-period takes place, voltage of negative polarity is applied to the winding of MagAmp TS1. During this time interval, MOSFET VT1 of the high-frequency push-

pull rectifier is in unconducting state (the voltage of negative polarity from the secondary winding  $W_{c1}$  is applied to its gate). Demagnetizing diode VD1 is conducting. Thus, the current flows through the control circuit 2, demagnetizing diode VD1, controlled MagAmp TS1, secondary winding  $W_2$  of high-frequency power transformer TV. This current is a function of the error signal obtained after comparison of regulator DC output voltage and reference voltage, and the change of the transformer secondary winding voltage due to the voltage change in the primary grid. The current causes demagnetization of MagAmp core from the saturation induction  $B_s$  to some induction  $B_1$ . The demagnetization depth is regulated with this stabilizing feedback. When the polarity of the input voltage changes to positive, remagnetization of the controlled MagAmp TS1 begins from the memorized level of induction  $B_1$ . When the controlled MagAmp TS1 operates in control half-period, the controlled MagAmp TS2 operates in a different mode – working half-period. During this time interval the rectifier MOSFET VT2 conducts (the voltage of positive polarity from the secondary winding  $W_{c2}$  is applied to its gate). The demagnetizing diode VD2 is not conducting. The current flows through secondary winding  $W_2$  of the power transformer TV, the winding of MagAmp TS2, rectifier MOSFET VT2, inductor L, capacitor C, load  $R_L$ . The working half-period consists of two subintervals. During the first subinterval the core of MagAmp TS2 remagnetizes from the memorized level of inductance to saturation inductance  $B_s$ . The time required for this remagnetization is considerably shorter than the demagnetization time of the control half-period due to no limitations of the remagnetization velocity (the load resistance is considerably smaller than the equivalent resistance of the control circuit). That's why the MagAmp core saturates within the half-period of the working frequency. During this time subinterval the current flows through inductor L, load  $R_L$ , and reverse diode VD3, the discharge current of the output LCD filter's capacitor C flows through the load  $R_L$  as well. After reaching saturation, the resistance of controlled MagAmp TS2 approaches zero, and the circuit current is defined with the load resistance  $R_L$  (second subinterval). Changing the depth of demagnetization of the controlled MagAmps TS1 and TS2 from  $+B_s$  to  $-B_s$  during the control half-period, we get the pulse-width modulation within a half-period of commutation high frequency during working half-period. This provides output voltage stabilization at change of the load current within its whole range. There also is a much lower level of losses in high-frequency rectifier due to significantly smaller resistance of channels of MOSFETs VT1, VT2 in conducting state, compared to the losses on diodes (when they are used in the rectifier) as a result of both direct voltage drop and, often, unsatisfactory frequency characteristics of diodes. Decrease of the equivalent resistance of the regulator allows obtaining higher level of load current along with high efficiency, if the operation modes of all topology components are agreed.

However, in such voltage regulator, the losses on the output filter reverse diode VD3 (about 1/3 of the load current flows through it) do not allow achieving maximum possible efficiency.

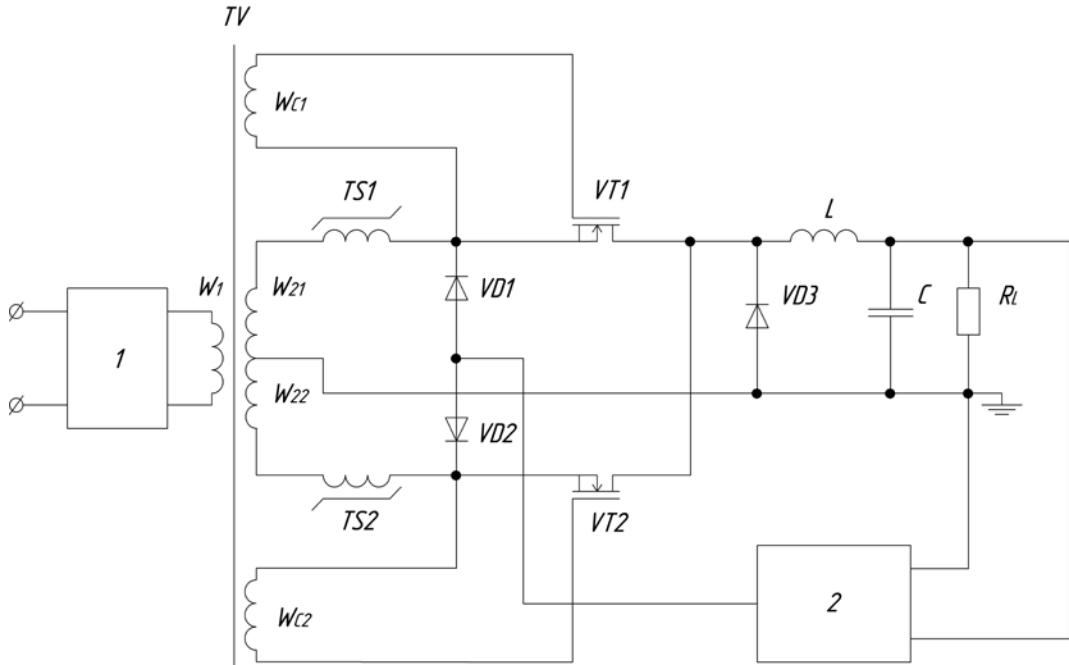


Fig. 4. The functional scheme of MagAmp power converter with synchronous rectifier

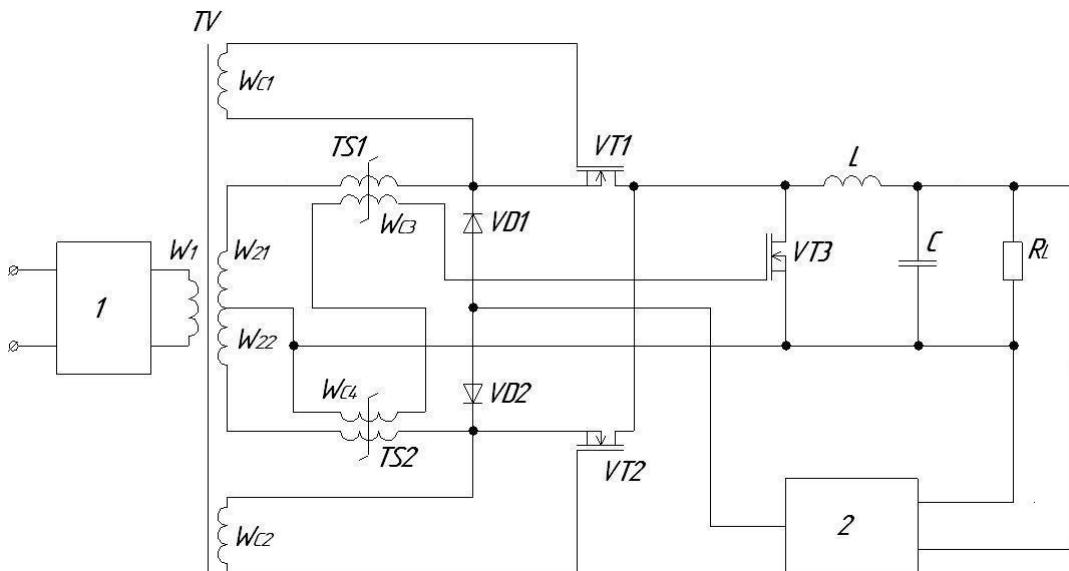


Fig. 5. The functional scheme of MagAmp power converter with synchronous rectifier with MOSFET instead of output LCD filter diode

It is suggested to substitute the output filter reverse diode with a MOSFET, that would be controlled in a function of voltages of additionally placed respective windings of controlled MagAmps.

The functional scheme of DC voltage regulator is shown in fig. 5 [13]. The waveforms that illustrate the principle of its operation are presented in fig. 6.

The advantage of using a synchronous rectifier in a power converter based on high-frequency MagAmps is that the load current starts flowing through it when its MOSFETs are already in conducting state. This is due to MagAmp operation principle. As a result, the converter's dynamic losses are decreased.

For instance, according to the experimental research, efficiency of the power converter based on high-frequency MagAmps with output parameters of 24V, 10A, where diodes have been used in the output rectifier and output filter, constituted 92% [14]. Its input active power was equal to 260.87 W. Which means the losses were equal to 20,87W, and about a half of those were the power losses in the output rectifier and filter. The use of MOSFETs with the open channel resistance of 0.2 mΩ in synchronous rectifier and output filter allows to significantly decrease these losses. The efficiency of such power converter is expected to be  $\geq 95\%$ . The efficiency tends to grow when designing power converters with higher output power.

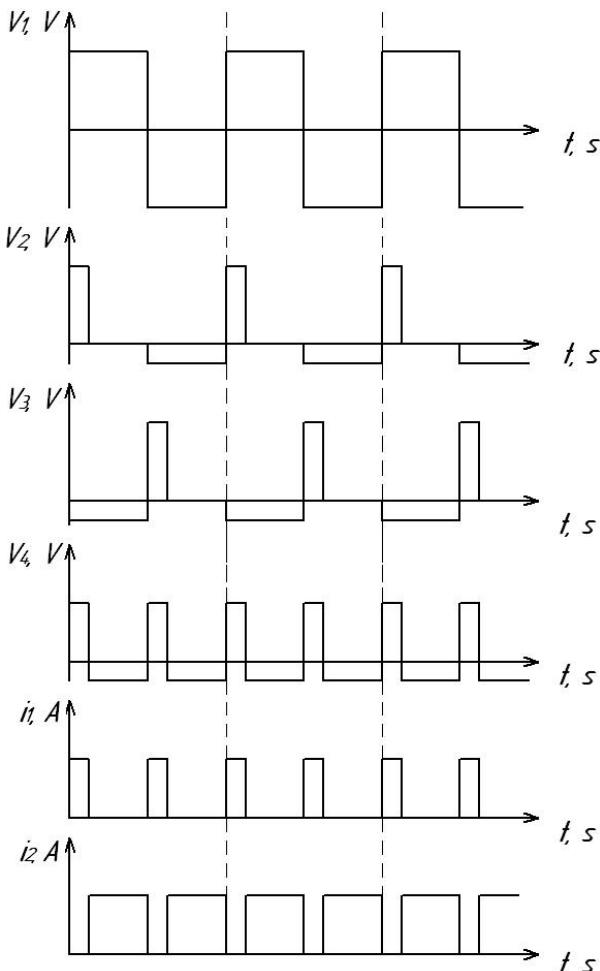


Fig. 6. Theoretical waveforms of MagAmp power converter with synchronous rectifier with MOSFET instead of output LCD filter diode

#### IV. CONCLUSION

Thus, the high level of load current along with high efficiency of the suggested DC voltage regulator are obtained due to:

- 1) the use of MOSFETs in the push-pull centre-tapped rectifier, which are synchronously controlled from the corresponding high-frequency power transformer secondary windings;
- 2) the use of MOSFET instead of reverse diode in output filter, which is controlled from the additionally placed corresponding windings of controlled MagAmps.

#### REFERENCES

- [1] L. Hua, J. Guo, X. Jing, S. Luo, "Design considerations for secondary side synchronous rectifier MOSFETs in phase shifted full bridge converters," in *Proceedings of Twenty Eighth Annual IEEE Applied Power Electronics Conference and Exposition (APEC)*, Long Beach, CA, USA, 17-21 March 2013, pp.526-531.
- [2] S. Mappu, "Control driven synchronous rectifiers in phase shifted full bridge converters," *Texas Instruments Application Note. Power Supply Control Products, SLUA287*, March 2003, 10 p.
- [3] B. Yang, J. Zhang, "Effect and utilization of common source inductance in synchronous rectification," in *Proceedings of Twentieth Annual IEEE Applied Power Electronics Conference and Exposition (APEC 2005)*, Austin, TX, USA, 6-10 March 2005, 5 p.
- [4] C. Fei, F. C. Lee, Q. Li, "Digital implementation of adaptive synchronous rectifier (SR) driving scheme for LLC resonant converters," in *Proceedings of Applied Power Electronics Conference and Exposition (APEC), IEEE 2016*, pp. 322-328.
- [5] M. S. Amouzandeh, B. Mahdavikah, A. Prodic, B. McDonald, "Digital synchronous rectification controller for LLC resonant converters," in *Proceedings of Applied Power Electronics Conference and Exposition (APEC), IEEE 2016*, pp. 329-333.
- [6] Y. Gu, Z. Lu, Z. Qian, G. Huang, "A novel driving scheme for synchronous rectifier suitable for modules in parallel," *IEEE Transactions on Power Electronics*, Vol. 20, No. 6, November 2005, pp. 1287-1293.
- [7] Linear Technology, "Synchronous rectifier driver for forward converters LTC3900," available online at: <http://www.analog.com/media/en/technical-documentation/data-sheets/3900fb.pdf>.
- [8] International Rectifier, "Automotive DirectFET power MOSFET AUIRF8736M2TR," January 14, 2014, available online at: <https://www.infineon.com/dgdl/aurif8736m2.pdf?fileId=546d462533600a4015355b0dade1414>.
- [9] MagAmp Cores and Materials, Technical Bulletin, BULLETIN SR-4, Magnetics Inc., available online at: <http://www.mag-inc.com/design/technical-documents>.
- [10] B. Mamano, "Magnetic amplifier control for simple, low-cost, secondary regulation," Unitrode corp. slup129, available online at: <http://www.ti.com/lit/ml/slup129/slup129.pdf>.
- [11] K. Harada, T. Nabeshima, "Applications of magnetic amplifiers to high-frequency dc-to-dc converters," *Proc. IEEE*, vol. 76, no. 4, April 1988. Pp. 355-361.
- [12] A. Yaskiv, V. Yaskiv, "DC voltage regulator," Patent of Ukraine № 112230, issue date 10.08.2016.
- [13] A. Yaskiv, V. Yaskiv, "DC voltage regulator," Patent of Ukraine № 112231, issue date 10.08.2016.
- [14] V. Yaskiv, A. Abramovitz, K. Smedley, A. Yaskiv, "MagAmp regulated isolated ac-dc converter with high power factor," *Special issue of journal COMMUNICATIONS - Scientific Letters of the University of Zilina*, ISSN 1335-4205, No. 1A/2015, pp. 28-34.

# Systems for Monitoring Modes and Disturbances in High-Voltage Transmission Lines

Yuriy Franko<sup>1</sup>, Julia Franko<sup>2</sup>

1. Department of Computer Technologies, Ternopil Volodymyr Hnatiuk National Pedagogical University, UKRAINE, Ternopil, 10 Vynnychenka str., email: franko@tnpu.edu.ua

2. Department of Computer Science, Ternopil National Economic University, UKRAINE, Ternopil, 8 Chehova str., email: julia.franko.j@gmail.com

**Abstract:** The article is devoted to the important practical task of using computerized systems of monitoring and control objects of electric power engineering, in which arise problems associated with the effectively identify abnormal situations in high-voltage electrical networks..The effectiveness of applying computerized monitoring systems to control electric power objects is substantiated. The structures and components of control and diagnostics systems for monitoring modes, recognition and authentication of disturbances in high-voltage transmission lines of 6-35 kV are studied. The "Strila" and "Altra's" functions are systematized as the network components of automated remote monitoring and controlling the technological equipment of electric substations. A sequence of operations for the recognition and authentication of load surge, earth fault and starting up powerful electric motors in high-voltage transmission lines is formalized. A structure and components of a microelectronic special-purpose processor protection relay of high-voltage transmission lines are presented.

**Keywords:** high-voltage transmission line, relay protection, disturbance, information monitoring systems, special-purpose processors.

## I. INTRODUCTION

One of the main tasks of electricity supply is the maintenance of equipment in an operational condition. Therefore, using new methods for diagnosing the technical condition using specialized microcontrollers and special processors with a high level of digital signal processing is one of the most actual tasks..The successful development of microprocessor and microelectronic technology has created favorable conditions for the development and large-scale replication of automated monitoring distributed computer systems (DCS) to monitor and control distant technological facilities in various industries. The examples of the succesful application of the SCADA systems of this class are the developments made by famous foreign firms and domestic enterprises, for example, Oven, Elektrosvit, Schneider, etc. [1-4]. The DCSs [5-6] designed and developed by Ternopil design bureau (TDB) "Strila" and the Institute of microprocessor control systems of electric power objects (IMCSPEO, Lviv) are being effectively introduced into the electric-power industry. These information systems are multifunctional, and embrace a wide spectrum of functions:

monitoring, authentication of and controlling the equipment of 6-35-kV electric substations.

The important condition for permanent expansion of the functional capabilities, improvement of the system characteristics, and perfection of the components is their being equiped with specialized microcontrollers and special-purpose digital signal processors. First of all, it concerns the recognition and authentication of disturbances in high-voltage transmission lines in case of load surge, short-term cable fault, earth fault and starting up powerful electric motors, for example, in the oil and gas industry. The theoretic base of the algorithms for digital signal processing of phase currents and voltages in case of transients and transmission line disturbances in is formed by the image recognition methods, as well as by the statistical, correlational, spectral, and entropy analysis.

## II. AUTOMATED REMOTE-CONTROL SYSTEM (ARCS) "STRILA"

The ARCS developed by the TDB "Strila" [5] is designed for remote control, collection of telesignallization and telemetering parameters at 6-35-kV electric substations.

A control point (CP) of the "Strila" telemetry executes previous signal processing of phase voltages and currents, archive registration of digital data, and transmission of coded messages to a distant dispatching point (DP) of the district electric grid (DEG). In addition, the CP receives and executes the commands remotely passed from the DP.

The basic structure of the operative-information complex "Strila", which belongs to the class of the SCADA systems, consists of two remote modules, which are installed at the substation of 35/10 kV, and a server hosted in the offices of dispatching services of DEG and Oblenergo.

The structure of the control point at the 35/10 kV substation is shown in Fig. 1.

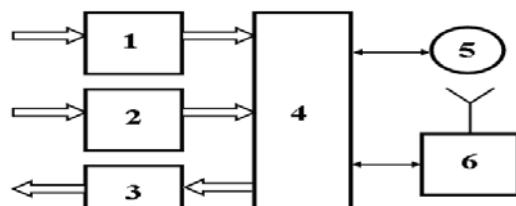


Fig. 1. Structure of CP : 1 - telesignallization module; 2 – telemonitoring module; 3 - telecontrol module; 4 - processor (microcontroller); 5 – operator's keyboard; 6 - radiomodem

The CP has the following information characteristics:

512 (32x16) discrete inputs (TS);

256 (16x16) analogue inputs (TV);

128 (16x8) control objects (TU);

RS-485 – master interface up to 2000 M;

MEK 870-5-101 – protocol of a code interaction between CP and DP.

While operating, the CP performs a component diagnostics and transmits information to DP. The structure of the "Strila" server equipment at DEG is shown in Fig. 2.

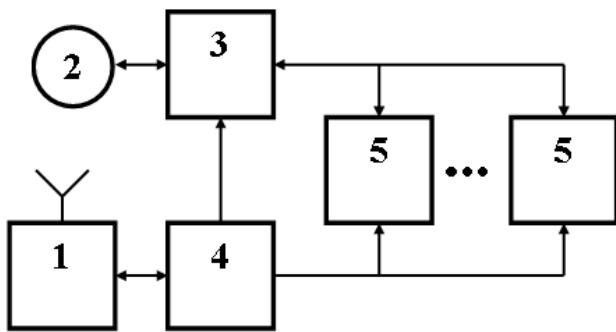


Fig. 2. Structure of server equipment "Strila" at DEG: 1 - radiodem; 2 - monitor of dispatch operator; 3 - safety system server; 4 - router; 5 - personal computers

The DP's server hardware of the complex "Strila" that is located at DEG executes:

3G CDMA EV-DO – safety system software;

IEC 60870-5-101 – computer network protocols;

IEC 60870-5-104 – integration with server equipment of the top-level electricity company CRM "UDS Consulting" (Call-centre) with a direct transfer of information on the position of energy facility switches.

The server system of DEG executes:

- graphic representation of tables of current, voltage and power of connections at the controlled electric substations;
- registration of system events (telecontrol, queries, program states, channel diagnostics, etc.);
- controlling the mimic diagram board, indexation of switch position and alarm system.

The described structure and functionality of the system "Strila" indicate that its level of microelectronic equipment and software meets the world's existing problem-oriented information systems of the SCADA class that are produced by leading foreign firms in the field of electricity.

At the same time, the classic reflection of the states of technological equipment at electric substations in the form of time trends, and tabular representation of digital data on the controlled parameters and processes greatly complicate the interaction of system operators, significantly reduce the speed and effectiveness of their responses to abnormalities of the states of technological facilities both before and at emergency. In addition, the system "Strila", as well as other systems of this class, is not intended to and does not carry out a deep statistical, correlational, spectral and entropy processing of telemetering signals with the possibility of recognition and authentication of disturbances in high-voltage transmission lines at invariant values of phase currents and voltages. For example, in case of load surge and

start-up of powerful electric drives, starting phase currents can considerably exceed those of earth-faults at a considerable distance from the substation's switching equipment with certain adjustments of relay protection facilities.

Such functions may be implemented successfully by specialized microelectronic processors on PLIS or by expansion of the problem orientation of application software of the computers embedded in such systems. In addition, the high performance of the special-purpose processors accelerating the signal processing by 1-2 orders of magnitude in comparison with the universal ones used in multifunctional microcontrollers allows the industrial equipment of power systems to be effectively and reliably secured on the intervals of 1-2 periods of industrial frequencies.

### III. CONTROL AND DIAGNOSTICS SYSTEM "ALTRA-MINI-ALTRA"

The system "Altra" developed by IMCSPEO [6] carries out local and distance monitoring of the insulation state of a 6-35 kV electric grid of arbitrary configuration with an insulated or compensated neutral. It performs a continuous real-time monitoring of the state of connections isolation at an electric substation, detection of a damaged area and its localization.

The system "Altra" performs the following functions:

- registration of partial insulation failure of arbitrary duration that provides the insulation state diagnostics of network segments;
- detection of an electric network segment with thinned insulation, remote transfer of information on its mode coordinates to the dispatching point;
- bay disconnection by single-phase earth fault protection, ensuring the appropriate selectivity depending on its kind, duration, category of consumers, features of insulants;
- displaying a network map on the monitor of the dispatching point, authentication of a thinned insulation segment;
- remote control of the substations' switching facilities by using the switchgears equipped with appropriate drives.

The system "Altra-Mini-Altra" is built on the multilevel principle where "Altra" and "Mini-Altra" are the basic blocks. "Altra" is installed at the substations and distribution points of 6-35 kV for the purpose of detection and registration of monophase earth faults. The device is designed for servicing up to 12 connections; the chart of its connections is shown in Fig.3.

Information for the "Altra" device is the phase voltages  $u_a$ ,  $u_b$ ,  $u_c$  and the value of the voltage of the zero sequence of  $3u_0$ , which are obtained from the sectional voltage transformer, as well as currents of zero sequence, which are obtained from current transformers of zero sequence of connections. Overall dimensions of the device are 40x30x22 cm. The device "Altra" can be installed on the facades of standard panels, as well as on its free parts or directly on the walls of the substation building or at the distribution substations. If necessary, to have information about the

position of the switching devices, the device "Altra" is supplemented by a block of control the state of discrete signals.

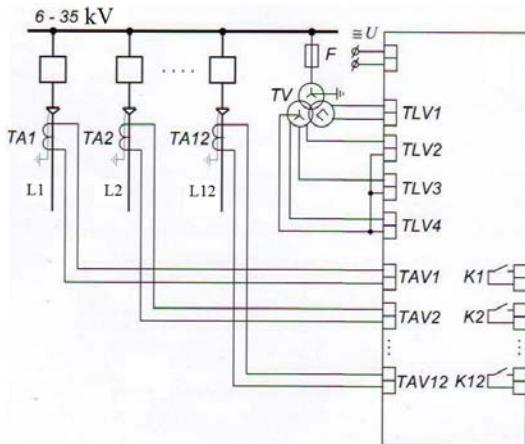


Fig. 3. Chart of connections and the "Altra" linking-up: TA – current transformer; TV – voltage transformer; L1 – L12 – connection lines; F – base fuse; K1 – K12 – switching facility

The block diagram of the "Altra" concentrator is shown in Fig. 4.

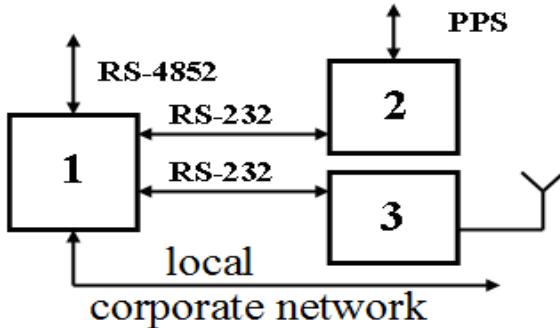


Fig. 4. Block diagram of the concentrator of the system "Altra": 1 – embedded computer; 2 – synchronization marks receiver; 3 – GSM-modem

"Altra" responds to the earthing of a connection phase. At any duration of the phase earthing, the date and time of the event when an insulation breakdown, and effective current of zero-phase sequence are detected can be seen on the "Altra's" display. Digital graphs of all mode coordinates are recorded in the power independent memory. The position of switching equipment is monitored by a block intended for monitoring the state of discrete signals. For the selective detection of a network segment with thinned insulation, the substations without voltage control are equipped with "Mini-Altra" devices (Fig. 5).

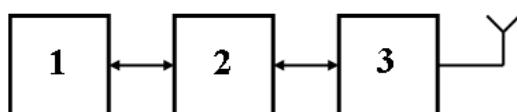


Fig. 5. Block diagram "Mini-Altra": 1 – processor; 2 – switching module; 3 – GSM-modem

This device controls the current of a zero sequence of substation connections (8 connections). The detection of a thinned insulation segment of the network with partly

voltage control at separate substations takes place on the basis of information received both from "Altra" and "Mini-Altra". Therefore, these devices are provided with 1 ms timing according to GPS-technology.

The data collection functions are performed by the concentrator based on the embedded computer.

The block diagram of data collection executed by the information system "Altra" is shown in Fig. 6.

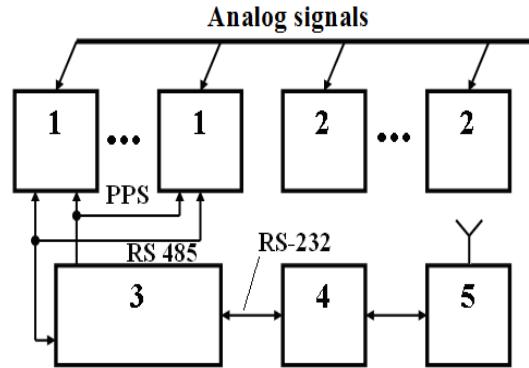


Fig. 6. Block diagram of data collection on the basis of the concentrator: 1 – Altra; 2 – Mini-Altra; 3 – concentrator; 4 – operator's workstation; 5 – GSM-modem

The improvement of the "Altra's" software and hardware, and the expansion of its functional capabilities is realized by reequipping the system with special-purpose processor protection relay of high-voltage transmission lines, whose structure is shown in Fig. 7 [6-9]. This is done on the basis of theoretical and experimental studies of transient processes in the transmission lines in case of charges and short circuits.

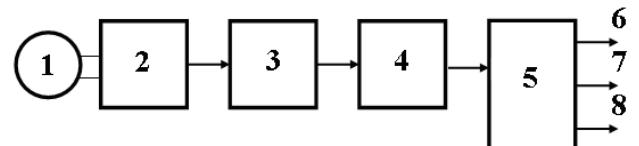


Fig. 7. Block diagram of a special-purpose processor device of relay protection: 1 - current transformer; 2 – two half period rectifier; 3 - analogue-to-digital converter; 4 – digital correlator; 5 – squarer; 6, 7, 8 - load surge signals, earth fault and starting up a powerful electric motor, respectively

The basis of the structural solution of this device is the task of improving the protection relay of high-voltage transmission lines by introducing a current rectifier, an analogue-to-digital converter, a logic element "Excluding OR", and a shift register. This makes it possible to recognize charges and short circuits in the transmission line, regardless of the absolute values of current growth in one of the phases.

The main advantages of the proposed relay protection device for high-voltage transmission lines are: advanced functionality; high speed recognition of charges and short circuits at intervals of one - two periods of industrial frequency; simplified algorithm of microelectronic implementation of the device; possibility of realization on a FPGA crystal; reduced cost of the device and the possibility of mass replication and implementation at high-voltage

substations; increased reliability and ability to work in a wide range of temperatures.

The device operation based on the method of integral-difference processing of phase currents in case of disturbances in high-voltage transmission lines developed by IMCSPEF in collaboration with the Department of specialized computer systems of Ternopil National Economic University [8,10].

Figure 8 shows an example of a change in the characteristics of digital graphs of a phase current at load surge (1), earth-faults (2) and starting up powerful electric drives (3).

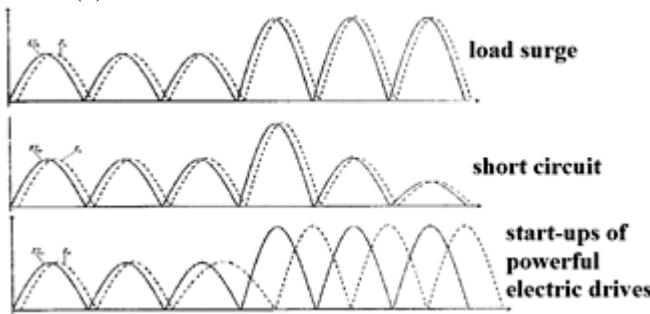


Fig. 8. Examples of disturbances in high-voltage transmission lines.

The results of computer modeling and identification of classified load surge (perturbations) in transmission lines are shown in Fig. 9, namely, integral characteristics of the modular differences in the squares of instantaneous current values of working and output currents, in the case of three-phase drives (the start of powerful electric motors). In particular, starting the powerful asynchronous motors, and long-term start-ups are shown in Fig. 9.

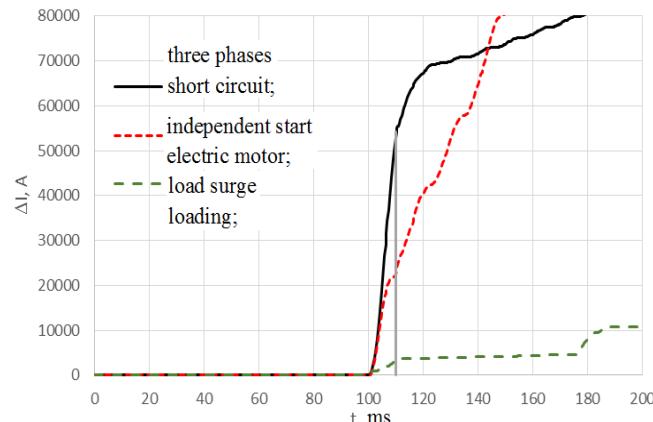


Fig. 9. Results of computer simulation of disturbances in transmission lines.

Fig. 9 shows that after the first period it is possible to judge which of the modes refers to a state of emergency and which is not. As the rapid linear change in the integral characteristics of the modular differences in the squares of instantaneous current values occurs during the first half-period, it is possible to recognize the emergency mode even during the half-period industrial frequency.

#### IV. CONCLUSION

The actuality and effectiveness of applying computerized systems of monitoring and control of electric power engineering objects is outlined. The functions and basic modules of the structures of "Strila" and "Altra" are systematized. The necessity of expanding functional capabilities of the systems of such class by using microelectronic special-purpose processors with a deep level processing of digital graphs of phase currents on the basis of statistical, cross-correlation, spectral and entropy analyses is substantiated.

The structure of the relay protection device based on the microelectronic special-purpose processor was designed and the results of computer simulation of the proposed integral-difference method of the correlation signal processing in case of disturbances in high-voltage transmission lines are presented. Such devices can widely be used both as the components of computerized systems for monitoring and control of the electric substations equipment and autonomous specialized protection relays. The control and diagnostic monitoring systems allow to effectively identifying abnormal situations in high-voltage electrical networks.

#### REFERENCES

- [1] Electronic resource: <https://www.abb.ua>
- [2] Electronic resource: <http://elektrosvit.com.ua>
- [3] Electronic resource: <https://owen.ua>
- [4] Electronic resource: <https://schneider-electric.ua>
- [5] Electronic resource: <http://tkbr-strila@ukr.net>
- [6] Electronic resource: <http://imskoe.org.ua>
- [7] Oleh Liura, Ivan Ostrovka, Iryna Sabadash, Yaroslav Nykolaichuk, "Theoretical Principles and Methods of Disturbances Recognition in Load Surges", in *Proc. Modern Problem of Radio Engineering, Telecommunications and Computer Science: TCSET'2016*, pp. 33–36, Slavskie in Lviv region, Ukraine, 2016.
- [8] Y.M. Nykolaichuk, N.Y. Vozna, O.P. Liura, I.I. Ostrovka, I.I. Sabadash, *Protection relay of high-voltage transmission lines*, Ukraine Patent № 103938, 2016, Bull. №1 (in Ukrainian).
- [9] O.P. Liura, N.Y. Vozna, "Research into and optimization of component characteristics of a microelectronic protection relay of high-voltage transmission lines", in *Proc. Scientific Bulletin of UNFU of Ukraine*, pp. 6–12, Vol. 27. № 5, Lviv, Ukraine 2017. (Ukrainian)
- [10] Tetiana Zavediuk, Yurii Franko, Oleg Liura, Ivan Ostrovka, "Methods of Recognition and Identification of disturbances in High-Voltage Power Lines", *proceedings of the XIII th International Conference CADSM 2015*, Lviv-Polyana, Lviv Polytechnic Publishing House, Ukraine, 2015, pp. 282–285.

# Artificial Intelligence Systems

# On Construction of Inductive Modeling Ontology as a Metamodel of the Subject Field

Halyna Pidnebesna, Volodymyr Stepashko

International Research and Training Centre for Information Technologies and Systems of the NAS and MES of Ukraine,  
Glushkov ave., 40, Kyiv, 03680, Ukraine, e-mails: stepashko@irtc.org.ua, pidnebesna@ukr.net

**Abstract:** The paper considers the constructing issue of ontology for the GMDH-based inductive modeling domain. It examines the main components of the GMDH algorithms in terms of their synthesis for designing the domain ontology to construct inductive modelling tools. Such approach significantly expands opportunities for construction of GMDH-based tools for building forecast models of complex processes of different nature.

**Keywords:** inductive modeling, GMDH, metamodel, domain ontology, model class, structure generator, selection criterion, computation tool.

## I. INTRODUCTION

With the development of technologies, the Internet, an extremely large and constantly growing number of information resources, there was the need to develop tools for automatic processing and analyzing data of different nature with taking into account the semantics (content) of this information. Some intelligent knowledge-based tools have been rapidly developed.

"Intelligent" computer systems (with artificial intelligence properties) may be understood as the ability to find ways to solve any task automatically, without (or with minimum) human intervention. The necessary features of such systems are adaptability, the ability to take into account the results obtained earlier, getting problem solutions by analogy with other cases, building valid (effective) algorithms, using the knowledge contained therein. That is, an intelligent computer system should be able to simulate the process of constructing an algorithm for solving the current problem, like human considerations.

In the design of computer modeling systems with artificial intelligence properties, the perspective direction of research is the use of an ontological approach to knowledge representation. This allows expanding computer capabilities, increases their "intelligence" and also simplifies the process of developing and modifying software products to solve specific tasks of constructing models and forecasts.

The advantage of an ontological approach is that ontology defines a conceptual structured environment in which the process of constructing a model of an object occurs [1]. This environment should be independent on the choice of a particular simulation object.

The automation task for intelligent computer modeling systems may be interpreted as modeling of the modeling process, which may be called as metamodeling. A metamodel is a model that describes the structure, principles of other models' operation.

The paper considers ways to increase efficiency of the

development of inductive modeling software.

The Group Method of Data Handling (GMDH) [2] is one of the most effective inductive modeling methods having intelligent properties [3]. It is the method of building models with automatic selection of structure and parameters on the basis of a short data sample with incomplete and uncertain input information to identify unknown relationships of the object or process under study.

In this paper, an approach is considered for "intellectualization" of inductive modeling software tools by applying an ontological approach to representing the knowledge of the subject field to design a knowledge base, computing tools and intelligent interface.

The aim of this research is to build ontology of the inductive modeling subject area. For this purpose, the analysis of the modeling process is done and main its stages are characterized. The results of the analysis and structuring of this area are presented. The basic components and characteristics of them are determined and main principles of the GMDH ontology construction are outlined.

## II. STRUCTURING KNOWLEDGE OF INDUCTIVE MODELING DOMAIN

Modeling is a process of studying a real object, in which only some of its specific characteristics, description, and conditional image are used. We consider the mathematical modeling that is studying the properties of an object by analyzing and constructing its mathematical model.

There are two main approaches to constructing mathematical models of objects: the theory-driven (or *deductive*) and the data-driven (or *inductive*) ones (Fig. 1).

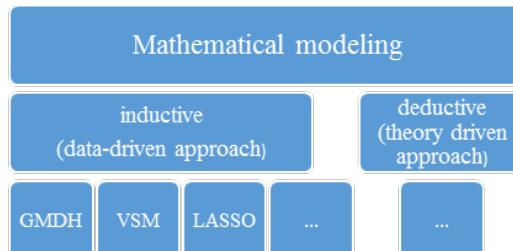


Fig. 1 Mathematical modelling approaches

Inductive modeling is the construction of a model based on the analysis and generalization of the statistical data about the object, obtained through observations or experiments.

Methods in the field include such algorithms for finding hidden patterns in data: GMDH, discovering of associative rules, sequence analysis, classification, regression, random forest, neural networks, support vector machine (SVM),

genetic algorithms, least absolute shrinkage and selection operator (LASSO) etc.

The inductive modeling algorithms solve a range of tasks:

- building mathematical models of objects/processes;
- forecasting processes specified by time series;
- construction of classification rules (supervised learning) for attributing an object to a given class;

- clustering (unsupervised learning or self-training; identification of effective features, forms and rules of distinction); in GMDH this problem is called “Objective Computer Clustering” (OCC);

- objective system analysis (OSA) when one need to find out which variables among the measured ones are independent (inputs), dependent (outputs) and irrelevant (uninformative) for building an appropriate model.

Inductive modeling based on statistical data is a process of sequential decision making, consisted of certain successive stages (Fig. 2). All methods of inductive modeling have standard components. This can be the basis of the metamodel of inductive modeling.

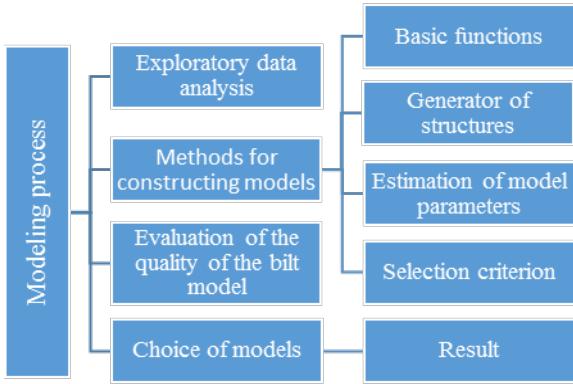


Fig.2 Components of inductive modeling process (fragment of the metamodel)

A metamodel provides the logical level of the domain and is interpreted dynamically at the application level. This adds additional flexibility to the system, since the domain logic can be changed without modifying the code. To allow or prohibit a particular type of communication at the logical level, it will suffice only to assign it to the formal terms of the metamodel.

In fact, the metamodel may be defined as a high level ontology, in terms of concepts of solution methods, key stages and constraints (Fig. 3). The ontological model of the subject domain of the lower level describes the algorithmic components of each particular modeling method in more details. To solve a practical task, an ontological model of a task is used having its own parameters, specific characteristics and areas of admissible values.

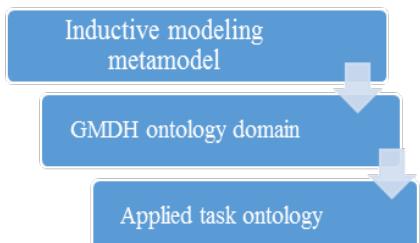


Fig.3 Hierarchy of GMDH domain ontologies

Any real problem can be characterized by the following main stages of the process of its solution: preparation; preliminary analysis; formulation of task; solving the task; analysis of results; their application. This upper level of structuring is supplemented by more detailed classifiers of subsequent hierarchical levels depending on the specificity of the problems under consideration.

The preparation of the task consists in determining the type of task (modeling of statics, time series or dynamics), modeling goals (approximation, interpolation, extrapolation, prediction, search for regularity), experiment planning (if the simulated system allows experimentation), obtaining a set of data (as a result of an active or passive experiment), their preliminary processing and organization of storage in the relevant database [4].

As a result of the field structuring, the principles of formation of algorithmic modules for solving a class of specific problem are determined. Depending on the type of tasks, adequate methods for solving them are selected.

Each of available methods corresponds to certain characteristics. According to them, it is possible to choose (may be automatically) a better method for a specific task. To do this, each of the set of solution methods should have some weight of importance to assess the adequacy of choosing this particular method at this stage. When choosing the appropriate (for a particular case) method at each stage of the modeling process, we get an algorithm (possibly the best) for solving a specific problem as a result of sequential synthesis in a structured set of possible options.

### III. METAMODEL AS THE HIGH LEVEL ONTOLOGY

To significantly expand the scope of computer modeling systems, they must be independent of the particular simulation object and of the means of its implementation. That means there should be a high level of abstraction of the subject area.

Raising the level of abstraction is difficult, and developers are forced to take out part of the information model for some application. This means fixing this part of the model. At the same time, the setting flexibility for the subject area is lost. The solution of this problem is seen in the introduction of metamodels. Metamodels reduce uncertainty in the description of the subject area and allow to get rid of rigid fixation on the task specificity.

First of all the metamodel helps to determine the structure of the process and allows developers to show specific requirements of the process automation means. The metamodel defines "design details", from which a modeling system may be subsequently created.

Metamodels are closely related to ontologies, because they are used to structure information and to analyze the relationships between concepts. The ontology divides the variables needed for some set of computations and establishes the relationship between them [5].

Modeling can be considered as an explicit description (design and rules) of how a problem-oriented model is constructed. As a rule, metamodels are a strict set of rules. A real metamodel is an ontology, but not all ontologies are represented explicitly as metamodels.

The internal structure of an intelligent computer system is

a reflection of certain knowledge that needs to be expressed explicitly, in a formal way. The use of ontologies can facilitate the description of the task of designing complex systems from components and implement a program that makes such a configuration independent of the product and the components itself, makes it possible to reuse.

*Ontology* is the exact specification of some field that contains a glossary of terms and a set of subject area links describing relations between these terms. It actually is a hierarchical conceptual skeleton of the subject area.

Formal ontology model ( $O$ ) is an ordered triplet [6]

$$O = \langle T, R, F \rangle,$$

where:

$T$  is finite set of terms of the subject area being described by the ontology  $O$ ;

$R$  is finite set of relations between the given terms;

$F$  is finite set of the interpretation functions given on the terms and/or relations of the ontology  $O$ .

The purpose of creating and using ontologies is support for activities to accumulate, distribute and reuse knowledge in a particular subject area.

Ontology allows one to specify a complex structure that can contain different types of data, provide a simple understanding of the presentation of structured knowledge and relatively easy updating.

In general case, the ontological model of the presentation contains a description of the situation/task (data, the purpose of the modeling) and the appropriate solution (algorithm for obtaining an adequate model). In most cases, in order to obtain an algorithm for solving a problem, it is sufficient a parametric representation in the form of a set of corresponding parameters given by the ontology, with specific values. In what follows, there is an example of an ontological representation of knowledge of the domain of inductive modeling.

#### IV. GMDH AS A METHOD OF MODEL BUILDING

The Group Method of Data Handling (GMDH) is one of the most effective methods of modeling from statistical data, which fully implements the essence of the inductive approach in modeling and has the intelligent properties.

GMDH is the method for constructing models with automatic determination of model structure and parameters from a data sample under conditions of incompleteness and uncertainty of input information in order to detect an unknown operation rule of an object or process under study.

GMDH characterizes by application of principles of automatic model generation with inductive complication of variants, non-definitive decisions and sequential selection according to external criteria for constructing models of optimal complexity. For comparison and selection of the best models, external criteria are used which are based on splitting the sample of input data into two or more parts. Estimation of parameters and quality assurance of models is carried out on different subsamples, which allows to automatically take into account different types of a priori uncertainty when constructing a model. These principles can be considered as metamodel characteristics for the process of building

mathematical model of an object (process).

#### V. ONTOLOGICAL MODEL OF GMDH-BASED INDUCTIVE MODELING PROCESS

To structure knowledge in a domain, one needs to consider the following issues:

- define the main stages for solving typical problems in a specific domain to obtain the basis for constructing the metamodel of the inductive modeling process;
- identify the main methods for effective solving these problems to form the basis of the domain ontology;
- generalize the experience of applying these methods to develop relevant intelligent software tools.

Obviously, each of these problems has a complex multilevel structure. The results of analysis of these problems are used to create the ontology of the subject field.

GMDH as one of the methods of inductive modeling also has a standard sequence of stages to solving a specific problem, as discussed in [7].

Ontology development is an integrated, sequential and iterative process. At the top level, the ontology contains a list of concepts and their general properties. In fact it is a thesaurus. A dictionary or a list of concepts is collected as a result of structuring knowledge domain. The next important step is to rank and organize the terms and build a hierarchy. In [8] a fragment of thesaurus of GMDH was given and general principles and main stages were described. The ideas given in [8] are substantially generalized in this paper.

The next step is more detailed study of GMDH algorithms, definition of the ontology structure and characteristics of the stages of the choice of a models class, structure generators, and model evaluation criteria. The result is the construction of corresponding ontological models.

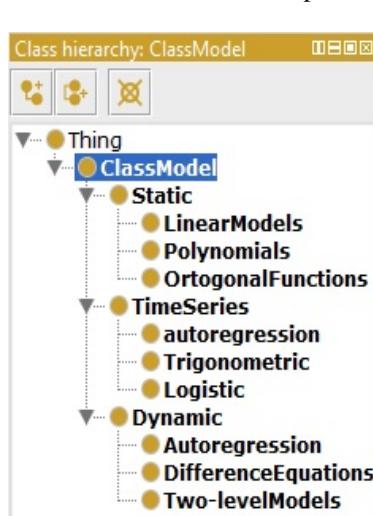


Fig. 4 The ontology of model classes

The ontology of classes of models CM (Fig. 4) is characterized by such key parameters as the number of input and output variables, and the number of past values (latencies) taken into account for input and output variables, respectively. Depending on the specific values of these parameters, one can obtain most of the variants of linear models that are used in practice to describe static objects, time series and dynamic objects and processes.

The *model generators* ontology GS (Fig. 5) contains two main types of GMDH structure generators: sorting-out and iterative ones. In turn, typical sorting-out algorithms to form different model structures are COMBI (exhaustive search) and multistage MULTI (directed search) [9]. Two main architectures of iterative structure generators are multilayer

## MIA and relaxational RIA.

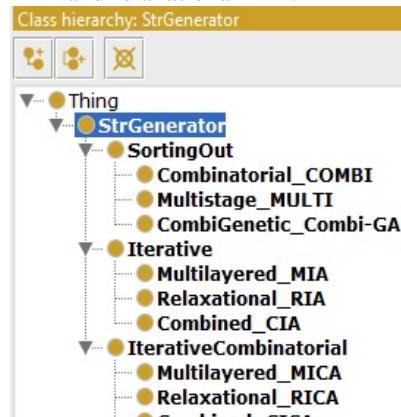


Fig. 5 The ontology of model generators

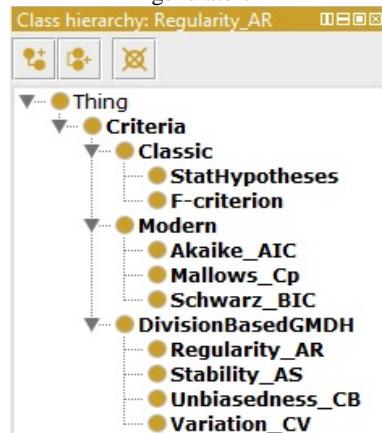


Fig. 6 Ontology of model criteria

In [8] an example is given for ontological model of sorting-out GMDH algorithm COMBI. The same way can be defining ontological models for other GMDH algorithms. For instance, let us consider the following set of parameters:

- an element of the model classes set  $CM$  is  $k_i^* = <\text{linear regression models}>$ ,
- an element of structure generators set  $GS$  is  $g_i^* = <\text{sorting-out algorithm:: directed search}>$ ,
- an element of parameter estimators set  $EP$  is  $p_i^* = <\text{least-squares method}>$ ,
- an element of selection criteria set  $CR$  is  $r_i^* = <\text{regularity criterion}>$ .

This set of parameters of the inductive modeling ontology defines the sorting-out GMDH algorithm MULTI [9].

In case when element of structure generators set  $GS$  is  $g_i^* = <\text{iterative algorithm:: relaxational}>$ , it defines the Relaxational iterative GMDH algorithm RIA.

In case element of structure generators set  $GS$  is  $g_i^* = <\text{iterative algorithm:: multilayered}>$ , it defines the Multilayered iterative GMDH algorithm MIA.

These are examples of ontology models as part of GMDH-based domain ontology. Preliminary analysis of the subject field enables the generalization of many different methods, identifying the key parameters. Ontology allows defining both general rules for constructing the algorithm and specific parameters when making an application.

In recent years, new kinds of GMDH algorithms have been developed: generalized iterative algorithm GIA [10] and the hybrid combinatorial-genetic algorithm Combi-GA [11]. So, taking into account current trends, the ontology of generators of structures may be presented in the form of Fig. 5.

Ontology of model criteria  $CR$  (Fig. 6) may be defined by key parameters that describe the penalty functions for the model complexity, the model quality, estimations of the unknown variance. They characterize a set  $CR$  of criteria, which are applied in practice for tasks of structural identification of models of optimal complexity.

## VI. CONCLUSION

The way to generalization of software tools of inductive modeling means by applying an ontological approach as metamodel representing the knowledge of GMDH-based domain is considered. This enables substantial simplification of developing specifications and software tools for solving various applied tasks.

The paper presents the results of structuring of the inductive modeling domain. The examples of main components of the modeling process defining their basic characteristics for building ontologies are provided. Some fragments of the ontology constructed using Protégé are presented as significant modules of the domain metamodel.

## REFERENCES

- [1] T. Gruber, "Toward principles for the design of ontologies used for knowledge sharing," *International Journal Human-Computer Studies*, 43(5-6), 1995, pp. 907-928
- [2] H.R. Madala, A.G. Ivakhnenko, *Inductive Learning Algorithms for Complex Systems Modeling*. New York: CRC Press, 1994, 384 p.
- [3] V.S. Stepashko, Conceptual fundamentals of intellectual modeling. *Control Systems and Computers*. – Kyiv: IRTC ITS, #4, pp. 3-15 (In Russian)
- [4] V.S. Stepashko, On the problem of structuring the expert's knowledge in the field of modeling by empirical data. ISSN 0454-9910. *Kibernetika i vychisl. tekhnika*. 1991, Issue 92, pp. 80-83. (in Russian)
- [5] Metamodelling, [cited 2017 Oct. 16]. Available from: <https://en.wikipedia.org/wiki/Metamodeling>.
- [6] T.A. Gavrilova, V.P. Khoroshevsky, *Knowledge Base Intelligent Systems*, SPb.: Piter, 2000, 384 p. (in Russian)
- [7] V. Stepashko, G. Pidnebesna, "Generalized Multifunctional Modules Concept for Construction of Inductive Modeling Tools," *Proc. of the 4th Int. Conf. on Inductive Modelling ICIM-2013*, Kyiv, Ukraine, Kyiv: IRTC ITS NASU, 2013, pp. 225-230.
- [8] H. Pidnebesna, On Constructing Ontology of the GMDH-based Inductive Modeling Domain, *Proc. of 8th International Workshop on Inductive Modeling IWM 2017*, Lviv, Ukraine, 2017, pp.511-513.
- [9] V.S. Stepashko, "A Finite Selection Procedure for Pruning an Exhaustive Search of Models," *Soviet Automatic Control*, 1983, vol. 16, no. 4, pp. 84-88.
- [10] V. Stepashko, O. Bulgakova, V. Zosimov Construction and Research of the Generalized Iterative GMDH Algorithm with Active Neurons. – In: *Advances in Intelligent Systems and Computing II. CSIT 2017 / Shakhovska N., Stepashko V. (eds). – Advances in Intelligent Systems and Computing, vol 689. Springer, Cham, 2018. – P. 492-510.*
- [11] V. Stepashko, O. Moroz, "Hybrid Searching GMDH-GA Algorithm for Solving Inductive Modeling Tasks," *IEEE Int. Conf. on Data Stream Mining & Processing*, Lviv, Ukraine, pp. 350-355, August 2016.

# Matrix Deep Neural Network and Its Rapid Learning in Data Science Tasks

Iryna Pliss<sup>1</sup>, Olena Boiko<sup>2</sup>, Valentyna Volkova<sup>3</sup>, Yevgeniy Bodyanskiy<sup>4</sup>

1. Control Systems Research Laboratory, Kharkiv National University of Radio Electronics, UKRAINE, Kharkiv, Nauky ave., 14,  
email: [iryna.pliss@nure.ua](mailto:iryna.pliss@nure.ua)
2. Control Systems Research Laboratory, Kharkiv National University of Radio Electronics, UKRAINE, Kharkiv, Nauky ave., 14,  
email: [olenaboiko@ukr.net](mailto:olenaboiko@ukr.net)
3. Samsung Electronics Ukraine Company, LLC R&D (SRK), UKRAINE, Kyiv, Lva Tolstogo St., 57,  
email: [v.volko@samsung.com](mailto:v.volko@samsung.com)
4. Control Systems Research Laboratory, Kharkiv National University of Radio Electronics, UKRAINE, Kharkiv, Nauky ave., 14,  
email: [yevgeniy.bodyanskiy@nure.ua](mailto:yevgeniy.bodyanskiy@nure.ua)

**Abstract:** The matrix deep neural network and its learning algorithm are proposed. This system allows reducing the number of tunable weights due to the rejection of the operations of vectorization-devectorization. It also saves the information between rows and columns of 2D inputs.

**Keywords:** deep learning, multilayer network, data mining, 2D network.

## I. INTRODUCTION

Nowadays, artificial neural networks (ANNs) are widely used to solve many problems arising in Data Science. Here, multilayer perceptron (MLP) [1,13-18] is the most widely used. On the basis of MLP deep neural networks (DNNs) [2-4,19,21] were developed, that have improved characteristics in comparison with their prototypes, namely traditional shallow neural networks.

In the general case, a multilayer perceptron that contains  $L$  information processing layers ( $L-1$  hidden layer and one output layer) realizes a nonlinear transformation that can be written in the form

$$\hat{Y}(k) = \Psi(X(k)) = \Psi^{[L]} \left( W^{[L]}(k-1) \Psi^{[L-1]} \times \right. \\ \left. \times \left( W^{[L-1]}(k-1) \Psi^{[L-2]} \left( \dots \Psi^{[1]} \left( W^{[1]}(k-1) X(k) \right) \right) \right) \right)$$

where:

- $\hat{Y}(k)$  denotes vector output signal of corresponding dimensions;
- $X(k)$  denotes vector input signal of corresponding dimensions;
- $\Psi^{[l]}$  are diagonal matrices of activation functions on each layer;
- $W^{[l]}(k-1)$  are matrices of synaptic weights that are adjusted during the learning process based on error backpropagation;
- $l = 1, 2, \dots, L$ ;
- $k = 1, 2, \dots$  is discrete time index.

In the DNN family, the most popular are the convolutional neural networks (CNNs) [20,22-25] that are mainly designed

to process images represented in the form of  $(n_1 \times n_2)$ -matrices  $X(k) = \{x_{i_1 i_2}(k)\}$  (where  $i_1 = 1, 2, \dots, n_1$  and  $i_2 = 1, 2, \dots, n_2$ ), which must be vectorized before submission to the network, i. e. they must be presented in the form of vectors [10], the dimension of which can be quite large, that leads to the effect of “curse of dimensionality”.

This effect can be avoided by processing the original matrix using convolution, pooling and encoding operations. As a result a vector of dimension smaller than  $(n_1 n_2 \times 1)$  is fed to the perceptron's input.

Although DNNs provide high quality of the information processing, their training time is too long, and the training process itself may require considerable computing resources. However, it is possible to speed up the information processing by bypassing the operations of vectorization-devectorization, i.e. by storing information that will be processed not in the form of a vector, but in the form of a matrix.

The abovementioned problem is solved by the matrix neural networks [5,6,11,12], that are quite complex from the computational point of view.

In this connection, it seems expedient to develop architecture and algorithms for tuning a deep matrix neural network that is characterized by the simplicity of the numerical realization and high speed of its synaptic weights learning.

## II. ADAPTIVE BILINEAR MODEL

The proposed matrix DNN is based on the adaptive matrix bilinear model introduced earlier by the authors [7, 8]

$$\hat{Y}(k) = \{\hat{y}_{j_1 j_2}\} = A(k-1) X(k) B(k-1), \\ j_1 = 1, 2, \dots, n_1; \\ j_2 = 1, 2, \dots, n_2 \quad (1)$$

where  $A(k-1)$ ,  $B(k-1)$  are  $(n_1 \times n_1)$ ,  $(n_2 \times n_2)$ -matrices of tunable parameters that are adjusted during online learning-identification process.

For this, either the gradient adaptation procedure

$$\begin{cases} A(k) = A(k-1) + \eta_A(k) \times \\ \quad \times E(k) B^T(k-1) X^T(k), \\ B(k) = B(k-1) + \eta_B(k) \times \\ \quad \times X^T(k-1) A^T(k) E_A(k) \end{cases} \quad (2)$$

is used or its version optimized by speed [7] that can be written as

$$\begin{cases} A(k) = A(k-1) + (Tr E(k) B^T(k-1) \times \\ \quad \times X^T(k) X(k) B(k-1) E^T(k)) \times \\ \quad \times (Tr E(k) B^T(k-1) X^T(k) X(k) \times \\ \quad \times B(k-1) B^T(k-1) X^T(k) X(k) \times \\ \quad \times B(k-1) E^T(k))^{-1} E(k) \times \\ \quad \times B^T(k-1) X^T(k), \\ B(k) = B(k-1) + (Tr E_A^T(k) A(k) X(k) \times \\ \quad \times X^T(k) A^T(k) E_A(k)) (Tr A(k) \times \\ \quad \times X(k) X^T(k) A^T(k) E_A(k) E_A^T(k) \times \\ \quad \times A(k) X(k) X^T(k) A^T(k))^{-1} \times \\ \quad \times X^T(k-1) A^T(k) E_A(k), \end{cases} \quad (3)$$

that is the matrix generalization of the Kaczmarz–Widrow–Hoff learning algorithm (here  $\eta_A(k)$ ,  $\eta_B(k)$  are learning rate parameters,

$$\begin{cases} E(k) = Y(k) - A(k-1) X(k) B(k-1), \\ E_A(k) = Y(k) - A(k) X(k) B(k-1), \end{cases}$$

$Y(k)$  is reference matrix signal).

The learning algorithm in Eq. (3) can be given additional filtering properties if the learning rate parameters in Eq. (2) are calculated using the recurrence relations that can be written in the form

$$\begin{aligned} \eta_A^{-1}(k) &= r_A(k) = \beta r_A(k-1) + \\ &+ Tr(E(k) B^T(k-1) \times \\ &\times X^T(k) X(k) B(k-1) \times \\ &\times B^T(k-1) X^T(k) X(k) \times \\ &\times B(k-1) E^T(k)) \end{aligned}$$

and

$$\begin{aligned} \eta_B^{-1}(k) &= r_B(k) = \beta r_B(k-1) + \\ &+ Tr(A(k) X(k) X^T(k) \times \\ &\times A^T(k) E_A(k) E_A^T(k) A(k) \times \\ &\times X(k) X^T(k) A^T(k)) \end{aligned}$$

where  $0 \leq \beta \leq 1$  is smoothing parameter [9].

On the basis of the model from Eq. (1), it is easy to introduce its nonlinear modification that can be written in the following form:

$$\begin{aligned} \hat{Y}(k) &= \left\{ \hat{y}_{j_1 j_2} \right\} = \Psi \odot (A(k-1) X(k) B(k-1)) = \\ &= \Psi \odot U(k), \end{aligned} \quad (4)$$

which is in fact the matrix generalization of the transformation that is realized by any of the layers of a multilayer perceptron.

In Eq. (4)  $\Psi$  denotes a  $(n_1 \times n_2)$ -matrix of activation functions, that acts elementwise on the matrix of internal activation signals of the system that are denoted by  $U(k) = \{u_{j_1 j_2}(k)\}$ .

In this case, the adjustment of the parameters of the nonlinear matrix model in Eq. (4) can be realized on the basis of the modified  $\delta$ -rule

$$\begin{cases} a_{j_1 j_2}(k) = a_{j_1 j_2}(k-1) + \eta_A(k) e_{j_1 j_2}(k) \times \\ \quad \times \psi'(u_{j_1 j_2}(k)) \sum_{i_2=1}^{n_2} b_{j_1 j_2}(k-1) x_{i_1 i_2}(k) = \\ = a_{j_1 j_2}(k-1) + \eta_A(k) e_{j_1 j_2}(k) \times \\ \quad \times \psi'(u_{j_1 j_2}(k)) \hat{x}_{i_1}(k) = a_{j_1 j_2}(k-1) + \\ + \eta_A(k) \delta_{j_1 j_2}(k) \hat{x}_{i_1}(k), \\ b_{j_1 j_2}(k) = b_{j_1 j_2}(k-1) + \eta_B(k) e_{A j_1 j_2}(k) \times \\ \quad \times \psi'(u_{A j_1 j_2}(k)) \sum_{i_1=1}^{n_1} a_{j_1 j_2}(k-1) x_{i_1 i_2}(k) = \\ = b_{j_1 j_2}(k-1) + \eta_B(k) e_{A j_1 j_2}(k) \times \\ \quad \times \psi'(u_{A j_1 j_2}(k)) \hat{x}_{i_2}(k) = b_{j_1 j_2}(k-1) + \\ + \eta_B(k) \delta_{A j_1 j_2}(k) \hat{x}_{i_2}(k). \end{cases} \quad (5)$$

On the basis of Eq. (4) it is easy to introduce into consideration a multilayer matrix neural network that realizes the transformation

$$\begin{aligned} \hat{Y}(k) &= \Psi \odot (A^{[L]}(k-1) (\Psi \odot (A^{[L-1]}(k-1) \times \\ &\times \dots \Psi \odot (A^{[1]}(k-1) X(k) B^{[1]}(k-1)) \dots) \times \\ &\times B^{[L-1]}(k-1)) B^{[L]}(k-1) \end{aligned} \quad (6)$$

Using the learning algorithm from Eq. (5) and error backpropagation, it is possible to obtain the adaptive procedure for tuning all parameters of the matrix DNN in Eq. (6):

- for the output layer:

$$\begin{cases} a_{j_1 j_2}^{[L]}(k) = a_{j_1 j_2}^{[L]}(k-1) + \eta_A(k) \delta_{j_1 j_2}^{[L]}(k) \hat{o}_{i_1}^{[L-1]}(k), \\ b_{j_1 j_2}^{[L]}(k) = b_{j_1 j_2}^{[L]}(k-1) + \eta_B(k) \delta_{A j_1 j_2}^{[L]}(k) \hat{o}_{A i_2}^{[L-1]}(k) \end{cases}$$

where

$$\begin{aligned} \hat{o}_{j_1 j_2}^{[L]}(k) &= \psi'(u_{j_1 j_2}^{[L]}(k)) e_{j_1 j_2}(k), \\ \hat{o}_{i_1}^{[L-1]}(k) &= \sum_{i_2=1}^{n_2} b_{j_1 j_2}^{[L]}(k-1) o_{i_1 i_2}^{[L-1]}(k), \\ \hat{o}_{A j_1 j_2}^{[L]}(k) &= \psi'(u_{A j_1 j_2}^{[L]}(k)) e_{A j_1 j_2}(k), \end{aligned}$$

$$\hat{o}_{A,i_2}^{[L-1]}(k) = \sum_{i_1=1}^{n_1} a_{j_1 j_2}^{[L]}(k) o_{i_1 i_2}^{[L-1]}(k);$$

- for the  $l$  th hidden layer,  $1 < l < L$  :

$$\begin{cases} a_{j_1 j_2}^{[l]}(k) = a_{j_1 j_2}^{[l]}(k-1) + \eta_A(k) \delta_{j_1 j_2}^{[l]}(k) \hat{o}_{i_1}^{[l-1]}(k), \\ b_{j_1 j_2}^{[l]}(k) = b_{j_1 j_2}^{[l]}(k-1) + \eta_B(k) \delta_{A,j_1 j_2}^{[l]}(k) \hat{o}_{A,i_2}^{[l-1]}(k) \end{cases}$$

where

$$\begin{aligned} \delta_{j_1 j_2}^{[l]}(k) &= \psi'(u_{j_1 j_2}^{[l]}(k)) \sum_{i_1=1}^{n_1} \delta_{j_1 j_2}^{[l+1]} a_{j_1 j_2}^{[l+1]}(k), \\ \hat{o}_{i_1}^{[l-1]}(k) &= \sum_{i_2=1}^{n_2} b_{j_1 j_2}^{[l]}(k-1) o_{i_1 i_2}^{[l-1]}(k), \\ \delta_{A,j_1 j_2}^{[l]}(k) &= \psi'(u_{A,j_1 j_2}^{[l]}(k)) \sum_{i_2=1}^{n_2} \delta_{A,j_1 j_2}^{[l+1]}(k) b_{j_1 j_2}^{[l+1]}(k), \\ \hat{o}_{A,i_2}^{[l-1]}(k) &= \sum_{i_1=1}^{n_1} a_{j_1 j_2}^{[l]}(k) o_{i_1 i_2}^{[l-1]}(k); \end{aligned}$$

- for the first hidden layer:

$$\begin{cases} a_{j_1 j_2}^{[1]}(k) = a_{j_1 j_2}^{[1]}(k-1) + \eta_A(k) \delta_{j_1 j_2}^{[1]}(k) \hat{o}_{i_1}^{[0]}(k), \\ b_{j_1 j_2}^{[1]}(k) = b_{j_1 j_2}^{[1]}(k-1) + \eta_B(k) \delta_{A,j_1 j_2}^{[1]}(k) \hat{o}_{A,i_2}^{[0]}(k) \end{cases}$$

where

$$\begin{aligned} \delta_{j_1 j_2}^{[1]}(k) &= \psi'(u_{j_1 j_2}^{[1]}(k)) \sum_{i_1=1}^{n_1} \delta_{j_1 j_2}^{[2]} a_{j_1 j_2}^{[2]}(k), \\ \hat{o}_{i_1}^{[0]}(k) &= \sum_{i_2=1}^{n_2} b_{j_1 j_2}^{[1]}(k-1) x_{i_1 i_2}(k), \\ \delta_{A,j_1 j_2}^{[1]}(k) &= \psi'(u_{A,j_1 j_2}^{[1]}(k)) \sum_{i_2=1}^{n_2} \delta_{A,j_1 j_2}^{[2]}(k) b_{j_1 j_2}^{[2]}(k), \\ \hat{o}_{A,i_2}^{[0]}(k) &= \sum_{i_1=1}^{n_1} a_{j_1 j_2}^{[1]}(k) x_{i_1 i_2}(k). \end{aligned}$$



Fig.1. Examples of the images from the MNIST dataset.

### III. COMPUTATIONAL EXPERIMENTS

The efficiency of the proposed system and learning methods was demonstrated on the classification task. A number of experiments was carried out on the MNIST dataset that was introduced by Yann LeCun and Corinna Cortes [26].

This dataset is widely used for training and testing in machine learning, namely in the classification task. This dataset contains 60000 training observations and 10000 test observations.

Each observation is an image of size 28x28 pixels that represents a handwritten digit. In general the dataset has 10 classes (digits from 0 to 9).

Some examples of the images from this dataset are presented in Fig. 1.

The elements of an image are represented by pixel values from 0 to 255, where 0 means white pixel (background) and 255 means black pixel (foreground). These values were preprocessed before training using normalization. The inputs for the network were  $(n_1 \times n_2)$ -matrices, where  $n_1 = n_2 = 28$ . Every hidden layer also had size of  $n_1 \times n_2 = 28 \times 28$ .

The results of the computational experiments are presented in Table 1.

TABLE 1. EXPERIMENTAL RESULTS

| Number of layers<br>in the network | Error on test set,<br>% |
|------------------------------------|-------------------------|
| 3                                  | 25                      |
| 5                                  | 20                      |
| 10                                 | 18                      |

#### IV. CONCLUSION

In this paper the matrix deep neural network and its learning algorithm are proposed. They allow significantly to reduce the number of adjustable weights due to the rejection of the vectorization-devectorization operations of 2D input signals.

One of the main advantages of the proposed system is that it also preserves the information between rows and columns of 2D inputs of the system.

The considered DNN in comparison with traditional multilayer perceptrons has increased speed, determined by reduced number of adjustable parameters and optimization of the learning algorithm, and the simplicity of numerical implementation.

The proposed system can be used to solve a wide range of machine learning tasks, particularly connected with the problems of image processing, where input signals are presented to the system for data processing in the form of a matrix.

#### REFERENCES

- [1] C. M. Bishop, *Neural Networks for Pattern Recognition*. Oxford : Clarendon Press, 1995.
- [2] Y. LeCun, Y. Bengio, G. Hinton, "Deep Learning," *Nature*, vol. 521, pp. 436-444, 2015.
- [3] J. Schmidhuber, "Deep Learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85-117, 2015.
- [4] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.
- [5] P. Daniušis and P. Vaitkus, "Neural networks with matrix inputs," *Informatica*, 19, №4, pp. 477-486, 2008.
- [6] J. Gao, Y. Guo, and Z. Wang, "Matrix neural networks," in *Proceedings of the 14th International Symposium on Neural Networks (ISNN), Part II*, Sapporo, Japan, 2017, pp. 1–10.
- [7] Ye. V. Bodyanskiy, I. P. Pliss, and V. A. Timofeev, "Discrete adaptive identification and extrapolation of two-dimensional fields," *Pattern Recognition and Image Analysis*, 5, №3, pp. 410-416, 1995.
- [8] S. Haykin, *Neural Networks: A Comprehensive Foundation*. Upper Saddle River, N. J. : Prentice Hall, Inc., 1999.
- [9] S. Vorobyov, Ye. Bodyanskiy, "On a non-parametric algorithm for smoothing parameter control in adaptive filtering," *Engineering Simulation*, vol. 16, p. 314-320, 1999.
- [10] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, "Deep learning for visual understanding: A review," *Neurocomputing*, vol. 187, pp. 27-48, 2016.
- [11] P. Stubberud, "A vector matrix real time backpropagation algorithm for recurrent neural networks that approximate multi-valued periodic functions," *International Journal on Computational Intelligence and Application*, 8(4), pp. 395-411, 2009.
- [12] M. Mohamadian, H. Afarideh, and F. Babapour, "New 2D Matrix-Based Neural Network for Image Processing Applications," *IAENG (International Association of Engineers) International Journal of Computer Science*, 42(3), pp. 265-274, 2015.
- [13] K. Suzuki, *Artificial Neural Networks: Architectures and Applications*. NY: InTech, 2013.
- [14] K. L. Du and M. Swamy, *Neural Networks and Statistical Learning*. Springer-Verlag London, 2014.
- [15] D. Graupe, *Principles of Artificial Neural Networks* (Advanced Series in Circuits and Systems). Singapore: World Scientific Publishing Co. Pte. Ltd., 2007.
- [16] L. Rutkowski, *Computational intelligence. Methods and techniques*, Berlin-Heidelberg: Springer-Verlag, 2008.
- [17] R. Kruse, C. Borgelt, F. Klawonn, C. Moewes, M. Steinbrecher, and P. Held, *Computational intelligence*, Berlin: Springer, 2013.
- [18] D. T. Pham and X. Liu, *Neural Networks for Identification, Prediction and Control*, London: Springer-Verlag, 1995.
- [19] I. Arel, D. Rose, and T. Karnowski, "Deep machine learning – a new frontier in artificial intelligence research," *IEEE Computational Intelligence Magazine*, vol. 5, no. 4, pp. 13-18, 2010.
- [20] K. Kavukcuoglu, P. Sermanet, Y-L. Boureau, K. Gregor, M. Mathieu, Y. LeCun, "Learning Convolutional Feature Hierarchies for Visual Recognition," in *Proceedings of the 23rd International Conference on Neural Information Processing Systems*, vol. 1, pp. 1090-1098, 2010.
- [21] C. Dan, U. Meier, and J. Schmidhuber, "Multi-column deep neural networks for image classification," in *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3642-3649, 2012.
- [22] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proceedings of the 25th International Conference on Neural Information Processing Systems (NIPS'12)*, vol. 1, pp. 1097-1105, 2012.
- [23] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770-778 2016.
- [24] Y. LeCun, K. Kavukcuoglu, and C. Farabet, "Convolutional networks and applications in vision," in *Proceedings of 2010 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 253-256, 2010.
- [25] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-9, 2015.
- [26] <http://yann.lecun.com/exdb/mnist/>

# Classification Model Based on Kohonen Maps

Jiří Jelínek

Institute of Applied Informatics, Faculty of Science, University of South Bohemia, CZECH REPUBLIC, Ceske Budejovice, 1760  
Branisovska St., email: [jelinek@prf.jcu.cz](mailto:jelinek@prf.jcu.cz)

**Abstract:** The standard Kohonen map uses unsupervised learning and single Kohonen layer, which allows the usage for clustering and visualization. The number of model parameters is relatively small and their settings are therefore not so complicated. The aim of this paper is to introduce three modifications of this basic model so that it can be used for classification tasks. The first change is the transition to supervised learning by adding input data about the required outputs. The second modification is the implementation of the hierarchical model structure to improve the classification results. The third extension is the implementation of an optimization mechanism for setting the parameters of the model because the number of model parameters was extended and their adjustment was more difficult. The results of the experiments with modified model will be presented too.

**Keywords:** Kohonen map, neural networks, pattern classification.

## I. INTRODUCTION

A variety of methods is available for advanced data processing, often from the field of artificial intelligence. Their use then depends on what data we have and what tasks we want to solve.

One group of possible tools for the implementation of machine learning or data analysis has quite a long been a neural network based on an artificial neuron model [4]. Probably the greatest attention was given to models of multilayered perceptron [5] and models based on this principle and using different methods of learning (e.g., back propagation of error). However, other models are also available (but the structure of the model described below, Kohonen maps, is similar to the multilayer network as well).

We use unsupervised and supervised learning methods for learning neural networks. When unsupervised learning is used, we only have unrated input data that we intend to analyze in some way. A typical example of unsupervised learning is the ART algorithm and model [1], which is capable to solve cluster analysis task.

When supervised learning is used, we train the model not only with the input patterns but also with the required output. These examples of the  $R^n > R^m$  transformation are used to form internal rules or model parameter settings. A typical example is a neural network with the algorithm of learning by error propagation [6].

The whole process of using the neural network model can be divided into two main phases - the setting phase (learning) and the production phase (recall). The production phase is the very reason for the existence of the model. In it the learned

settings are used for processing of previously unseen (test) input data.

A very interesting model is the so-called Kohonen map primarily designed for unsupervised learning and therefore cluster data analysis and visualization. However, the results of its use by the author on previously solved tasks [2] revealed that the use of the standard model and learning did not always lead to the desired results and it was necessary to solve also the classification tasks with a predetermined classification of inputs in addition to the clustering.

Therefore, a modified learning algorithm and a multi-level model structure based on Kohonen maps were designed. The very basic model was at first used for economic data processing [7]. The aim of this paper is to present actual state of the model with key changes in the hierarchical learning and recall and also in modifying the learning process. These changes were supposed to improve the quality of the model and its generalization capability, which was tested in experiments as well.

The following chapters of the paper are organized as follows. Chapter 2 focuses on a brief description of the standard Kohonen map model and shows the key parameters of it. Chapter 3 then concentrates on the description of the modifications that were made, and Chapter 4 focuses on experiments with the model conducted primarily to verify the benefits of the proposed modifications.

## II. RELATED WORK

The Kohonen map [3] was first introduced in the 1980s as most models of neural networks of different types. In some ways, Kohonen map can remind us of the ART2 model [1]. The similarity lies in the same requirements on input data (number vectors). Similar is also the two-layer structure of the network. However, Kohonen map is strongly focused on the visual interpretation of the output and is therefore useful both for the better understanding of the task and for use in an online dynamic environment. An example of such usage can be monitoring of the state of the system [2].

The core model activity is basically the same as ART2 - assigning input patterns to the cells of the second layer (i.e. the output clusters representing by these cells) based on the similarity of patterns.

### Basic Model Functionality

The input layer of the Kohonen map is composed of the same number of cells as the dimension  $n$  of the input space  $R^n$  is. The output layer is two-dimensional and is also referred to as the Kohonen layer. The input and output layers are fully interconnected from the input to the output one with links whose weights are interpretable as the centroid of the input

patterns cluster represented by the cell of the output layer. The number of output layer cells is the model parameter.

In the production phase, test patterns are submitted at the input of the model. Their distance (here the Euclidean one) is calculated from the output layer cell centroids and the input pattern is assigned to the output cell, from which it has the smallest distance. Assigning the input pattern to the output cell can be visualized in the output layer as shown in Fig. 1.

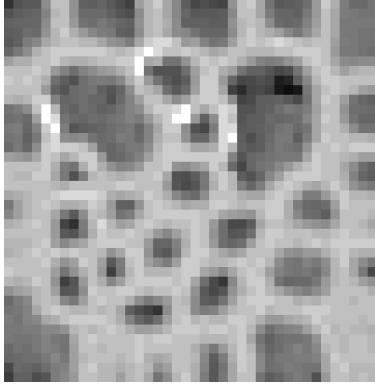


Fig. 1. Kohonen layer visualized according to the number of patterns represented by cells (the more patterns, the darker color).

The learning of the Kohonen map is iterative and is the extension of the production phase. During it, the values of the weights leading to the cell representing the pattern are modified (the winning cell, the pattern was assigned to it) so as the weights to the cells in its neighborhood. The centroids of all these cells are moved towards the vector representing the pattern according to the Eq. (1).

$$c_i^{new} = (1 - \alpha)c_i^{old} + \alpha s_i \quad (1)$$

In it the  $\alpha$  is the learning coefficient for the winning cell, usually with the value from the interval  $(0,1)$ ,  $c_i$  is the  $i$ -th coordinate of the cell's centroid and  $s_i$  is the  $i$ -th coordinate of the given input pattern.

The values of weights leading to the neighbor cells are modified according to the same formula, but with a different  $\alpha_{ij}$  learning coefficient value adjusted to respect the distance of a particular cell from the winning one:

$$\alpha_{ij} = \frac{\alpha}{(1 + d_{ij})} \quad (2)$$

Coordinates are taken relative to the winning cell with position  $[0, 0]$ . The distance  $d_{ij}$  from the winning cell in these relative coordinates is then calculated as Euclidean one. The neighborhood of the winning cell is then defined by the limit value of  $d_{ij} \leq d_{max}$ .

The Kohonen map also includes a mechanism to equalize the frequency of cell victory in the output layer. For each of them, the normalized frequency of their victories in the representation of the training models  $f_q$  is calculated with the normalized value in the interval  $(0, 1)$ . This is then used to modify the distance of the pattern from the centroid.

$$w_{pq} = d_{pq}(f_q + K) \quad (3)$$

In the formula,  $w_{pq}$  is the modified pattern distance of pattern  $p$  from centroid  $q$ ,  $d_{pq}$  the original Euclidean distance of them

and  $K$  global model parameter to limit the effect of the equalization mechanism. The calculated distance  $w_{pq}$  is then used in the learning process. The described mechanism ensures that during the learning the weights of the whole Kohonen layer will gradually be adjusted. The size of this layer together with the number  $P$  of input patterns actually determines the sensitivity of the network to the differences between the input patterns.

It was also necessary to choose the appropriate criterion for determining the end of learning. This is based on the average normalized distance  $v$  in the 2D layer through which the pattern "shifts" between the two iterations as shown in the Eq. (4).

$$v = \frac{1}{\sqrt{2NP}} \sum_p \sqrt{\Delta i^2 + \Delta j^2} \quad (4)$$

The distance is calculated on the square-shaped Kohonen layer (with  $N$  cells on the side) between the two iterations of the input set  $P$ . The  $v$  value is compared to the maximum allowable value  $v_{max}$  determining the average maximum shift of patterns allowed in one iteration.

The main parameters of the Kohonen maps are the learning coefficient  $\alpha$ , the way of setting the decrease of this coefficient for the cells around the winning cell, the size of the neighborhood given by  $d_{max}$  and the number of cells in the two-dimensional output layer. In addition, the behavior of the model also influences the value  $v_{max}$  and the coefficient  $K$  in Eq. (3) and its possible change over time.

### III. MODEL MODIFICATIONS

The modified learning algorithm and the multilevel structure of the model were designed together with optimization mechanism for model settings.

#### *Modified Learning*

The goal of learning is to get an adjusted model that is capable of certain generalization, i.e. adequate responses to patterns that have not been trained. The basic Kohonen model described above used unsupervised learning. The modification using supervised learning for solving classification tasks will be introduced.

At the learning stage, the first change was in the fact that after finding the winning cell in the Kohonen layer, the model also stores the classification of the training patterns that were assigned to it. With the help of a known classification of the training set, we can determine the probability rate of output categories in every cell, which can be used in the production phase to evaluate the test patterns in one of these two ways:

- *Probability evaluation.* We select the most likely category, so that the test pattern is always classified. The assumption here is the same distribution of *a priori* probabilities of output categories.
- *Absolute evaluation.* We only evaluate a test pattern if it is assigned to a winning cell representing only one category. A significant percentage of patterns can thus be unclassified.

So the modified model already uses at the production stage additional output information.

In the course of model experiments it has also been shown,

that classification task on data with complex transformation  $R^n > R^m$  tend to a state where the same classified patterns are assigned to output layer cells that are often very distant from each other. This affects the overall efficiency of the model that must respect this fragmentation.

The aim was to limit this phenomenon by using the output categorization directly in the model's learning phase. In this case, the model is trained on data that are a conjunction of the original input and the desired output (classification). For example, if we have a classification task performing the transformation  $R^4 > B^1$ , where  $B^1$  represents a one-dimensional binary space (one binary coordinate), the model will be taught on the input set  $R^4 \cap B^1$ . In the production phase, the last coordinate  $b_1$  is not used in the calculations because the input test vectors will not contain it (their classification is not known).

The described modification significantly changes the model settings, but it has turned out to be a positive change under certain conditions. The key is to what extent the output (often binary) classification should be projected into the training input. If this projection is in full binary value and inputs from  $R^4$  are normalized to a range (0; 1), the model settings are distorted too much and model is not capable to generalize. Therefore, the new reduction factor  $u$  was implemented to limit this projection according to Eq. (5)

$$inp_{x+|R|} = ub_x, \quad (5)$$

where  $inp_{x+|R|}$  is the value of the input of the model (preceded by the coordinates of the original input from the  $R$  space) and  $b_x$  is the original classification ( $b_x = 0$  or  $b_x = 1$  for pure binary classification). The factor  $u$  has a value from the interval (0; 1) and represents next parameter of the model.

#### Hierarchy Structure

The fundamental change in the work with Kohonen map is its repeated use with a different training set. This set can be e. g. quite uneven in terms of the representation of output categories or too large for the actual size of Kohonen layer. The modified model addresses this problem by gradually reducing this set by eliminating properly classified training patterns at the end of each learning iteration. In the next step, a new instance of the map is already learned with a training set containing only problematic (not yet categorized) patterns. The underlying idea of this approach is to use Kohonen map internal mechanisms so that the map in every step refines its classification capabilities.

Thus, in each model step, a separate Kohonen map is used. After learning, it is examined whether only patterns of one output category are assigned to the given cell. If this is the case, we can say that the map can correctly classify these patterns in accordance with the desired output and they can be excluded from the training set (Fig. 2). The successful classification of input pattern is considered as:

- *Basic classification.* Assignment to a cell representing only patterns of the same category.
- *Strict classification.* Assignment of the pattern to a cell according to the previous bullet and additionally adjacent to only the same (representing the same category) or empty cells (not representing any pattern).

The selection of one of the above classification methods is a parameter of the model.

Two criteria are crucial for the real use of the proposed model. The first one is the criterion of learning termination in each step (level) of hierarchical model learning. The criterion of the maximum average shift distance between the 2D layer cells according Eq. (4) was used.

The second criterion is that of the overall ability of the whole set of learned sub-models to correctly classify the training and later the test set of patterns. In the model was used the minimum size of the training set, which still makes sense for learning. Its higher value reduces the number of hierarchical classification steps but also limits the sensitivity of the model.

In the production phase, the classification method differs depending on whether we are in the last hierarchical step or not. For the last model in the structure, the probability evaluation is always used, where the pattern is assigned to the most likely category resulted from the learning process.

The modified model was set up using 11 parameters including both the Kohonen map original ones (used in every iteration) and the other ones characterizing the hierarchical model's operation.

## IV. EXPERIMENTS

The experiments carried out were aimed at confirming the preliminary hypothesis that both modifications of the Kohonen map model are beneficial to the generalization quality and hence the classification of the test set of patterns. The test data were artificially created to represent a complex nonlinear transformation from the input space  $R^4$  to the space  $B^1$ . 10,000 training and test patterns were generated with random coordinate values in the interval (0; 1) from  $R^4$  space.

One test set and two training sets were created. From the training sets one was for the classical learning of the model (only inputs from  $R^4$ ) and the other one extended with the output  $b_1$  (the network input dimension increased to  $R^5$  where the fifth coordinate was created from  $b_1$ ). Experiments have been optimized for maximizing the number of properly classified test patterns.

As mentioned above, the model has a number of adjustable parameters that significantly affect its results. Searching for optimal setup manually would be a lengthy process, and therefore, a superstructure of the model was used implementing an optimization mechanism based on genetic algorithms. The possible values (for categorical data) or limit minimal and maximal values (for number data) were inserted into the optimization system and the optimal values were searched through crossings and mutations of coordinates of settings chromosome. The chromosome evaluation function was defined as maximum correctly classified input vectors.

Two variants of model's setting have been examined. The first one focused on modified model learning and worked with a single level of the hierarchical model (as if the model was not hierarchically modified). The second variant involved a hierarchical modification of the model structure with the maximum number of levels limited to 10.

The best results are presented in Table 1. These are the best from 406 experimental settings of the model calculated by the

genetic algorithm for each variant.

It is clear from the table that the use of the modified learning process brings a significant improvement in the classification capabilities of the network. The key output is the finding that to achieve better results with normalized data, the output values used for learning (originally 0 and 1) must be reduced by the reduction factor. Its appropriate setting was found by genetic algorithm and was 0.2 or 0.3 (see the Reduction factor row in Table 1).

TABLE 1. BEST RESULTS FOR DIFFERENT MODEL SETTINGS

| Parameter                          | Max. 1 level |          | Max. 10 levels |          |
|------------------------------------|--------------|----------|----------------|----------|
| Input                              | classic      | extended | classic        | extended |
| Number of 2D cells                 | 40*40        | 35*35    | 40*40          | 35*35    |
| Classification criterion           | basic        | strict   | strict         | strict   |
| Reduction factor                   | -            | 0.2      | -              | 0.3      |
| Real used levels                   | 1            | 1        | 10             | 3        |
| Total learning iterations          | 33           | 118      | 238            | 515      |
| Test patterns correctly classified | 7998         | 8472     | 7985           | 8643     |
| Model success [%]                  | 79.98        | 84.72    | 79.85          | 86.43    |

The visual outputs of the 2D network for the setting from the last two columns of Table 1 in level 1 are shown in Fig. 2 to demonstrate the effect of modified learning. The cells representing the patterns rated 0 are green, the patterns rated 1 are red. We get "clean" colors for cells containing input patterns included in only one category, for cells containing patterns of different categories the color is mixed. This respects the number of patterns in a cell with different output categories. The influence of additional output information on the final network setting is quite obvious (right) and the fragmentation when using the classical learning algorithm (left) almost did not occur.

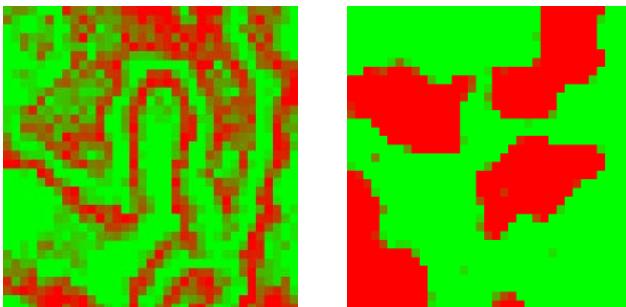


Fig. 2. Influence of modified learning algorithm.

The results can still be improved by using a hierarchical modification of the model, but for the selected transformation  $R^n > R^m$  the added value is not so high (quality improvement 1.71 %). In this case, the training set was evenly generated, but the benefit will be more significant on data with

unequally represented output categories or different *a priori* probabilities of them.

## V. CONCLUSION

This paper focuses on introducing a modified learning algorithm for the Kohonen map, which enables solving of the classification tasks. The core of the modification is the use of training set extended with the output categorization of the training patterns. Modifications were made even in setting of the criteria for completing model learning (criterion of minimal pattern shift in the 2D layer). A visual superstructure of the model was also developed to allow a detailed study of the dynamics of the model setup process, and the obtained knowledge could be used to better understanding of the learning process and the nature of input data.

The second important modification is the design and description of the behavior of a hierarchical classification model based on modified Kohonen maps. The training set is gradually reduced during the process of learning. This increases the sensitivity of the network to differences in input data. The algorithm for the production phase of the model was developed, based on the learned Kohonen map sub-models.

The behavior of the hierarchical model was described by a series of input parameters, whose values had to be empirically determined. Therefore, to find optimal values, optimization system based on genetic algorithms has been used.

With the modified model, experiments were conducted to verify the benefits of the proposed modifications. They confirmed the positive influence of the extended training set and the hierarchical structure of the model for better classification performance. The overall classification quality was improved by 6.58% on the generated data with nonlinear randomly selected transformation function  $R^4 > B^l$ . The benefit of the hierarchical structure would be greater when using data unevenly covering the input space.

Future work on the model will focus on further examining the benefits of proposed modifications to the quality of the classification process. Attention will also be paid to an optimization mechanism that could include more data characterizing the model's activity.

## REFERENCES

- [1] G. A. Carpenter and S. Grossberg, "ART 2: Self-organization of stable category recognition codes for analog input patterns", in *Applied optics*, Vol. 26(23), 1987, pp. 4919-4930.
- [2] J. Jelínek, *Exceptional working states of power grids*. Qualification work. CTU - FEE, Prague, 1992. (in Czech)
- [3] T. Kohonen, "Self-organized formation of topologically correct feature maps", in *Biological cybernetics*, Vol. 43(1), 1982, pp. 59-69.
- [4] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity", in *The bulletin of mathematical biophysics*, Vol. 5(4), 1943, pp. 115-133.
- [5] F. Rosenblatt, *The perceptron, a perceiving and recognizing automaton Project Para*. Cornell Aeronautical Labs., 1952.
- [6] D. E. Rumelhart, G. E. Hinton and R. J. Williams, "Learning representations by back-propagating errors", in *Nature*, Vol. 323, 1986.
- [7] M. Vochozka, J. Jelínek, J. Váchal, J. Straková and V. Stehel, *Using of neural networks for comprehensive business evaluation*. H. C. Beck, Prague, 2017. (in Czech).

# Reasoning With Streamed Information from Unreliable Sources

Ladislav Beránek, Radim Remeš

Department of Applied Mathematics and Informatics, Faculty of Economics, University of South Bohemia, CZECH REPUBLIC, Ceske Budejovice, 13 Studentska str., email: beranek@ef.jcu.cz, inrem@ef.jcu.cz

**Abstract:** At present, we are increasingly struggling with the need to make decisions based on information from data streams from different sources which are often unreliable. When deciding, we need to process the this observed information, and we must estimate their reliability. In this paper, we propose a framework that allows us to derive information from unreliable sources and to estimate their trustworthy. This framework is fully implemented on data streams with the aim to derive of new facts from incoming information. This information is coming as unstructured messages that are transmitted from heterogeneous and potentially untrustworthy sources. This information is processed using a natural language and belief function theory. The trustworthy of processed information is estimated based on their internal conflict. The proposed framework is evaluated using an experiment that quantifies the efficiency of our solution with respect to accuracy and overhead of the proposed framework.

**Keywords:** streamed data, reasoning, belief function theory, uncertainty.

## I. INTRODUCTION

Users share and process all sorts of data within various applications based on the Internet infrastructure. Users evaluate various products and express their opinions to different events. Tripadvisor.com server might be an example. Users can evaluate here certain hotel on the base of their satisfaction with its services. Wikipedia provides feedback tool to engage the reader in a review of article quality based on four criteria, i.e., "trustworthy", "objective", "complete" and "well written". Such activity is referred also as crowdsourcing. Many users are used for evaluation or classification of certain product or services. It is sometimes used in science. An example might be the website Galaxy Zoo, where users classify astronomical images.

As this method is useful, organizers usually have little control over quality of users' activity. Reaction of individual users may vary substantially, and in some cases, they may even be controversial. The question is then how to integrate feedback from multiple users to get an objective opinion. Commonly used heuristics such as "majority voting" and "take the average" ignore individual user experience and can fail, for example in an environment where there are users with malicious intent. The aim of this paper is to propose and to test a method to determine the grand truth without knowing the previous experience of users. For this purpose, it is used an approach based on the Dempster-Shafer theory.

Within this theory, the operation discounting is defined. At

this operation, the value of belief function varies in dependence on certain additional information or if the pieces of information, to be integrated, are contradictory. When it is necessary to decide to implement discounting process the following questions are to be solved: What resources are to be discounted? Up to what extent these resources should be discounted? The model used in this paper introduces an iterative method which automatically determines the discount rate on the base of the reliability of sources. The advantage of this approach is that it does not require any additional meta-information about the reliability of sources. The method assumes only that the more specific source of information conflicts with the majority opinion, the stronger this source must be discounted.

The rest of this paper is organized as follows. Section 2 provides an overview of related work. Section 3 formulates the problem and introduces a belief function framework with the proposed model. Section 4 presents experimental results on synthetic data. Conclusions are composed in Section 5.

## II. RELATED WORK

Currently, several studies deal with the setting involving multiple labelers. For example, the work such as [4, 7, 9, 19, 20, 22] focus on the estimating the error rates of observers. Authors [4] deal with selecting the best set of all available information from users for model training. These works focus on learning classifiers directly from user data instead of estimating ground truth. Work [14, 15] uses a probabilistic framework for solving classification, regression and ordinal regression problem with multiple annotators. This framework assumes that the expertise of each annotator does not depend on these data. Works [23, 25, 26] develop this approach, but do not build fully on this premise. There are some other related works, which focuses on a different setting [3, 24]. Recent work [8] pays attention to regression problem under multiple observers, with the use of less parametric methods for modeling and designing observers regression function.

## III. METHODOLOGY

### *Belief function theory framework*

Our model is an application of the Dempster-Shafer theory. The Dempster-Shafer theory [16] is designed to deal with the uncertainty and incompleteness of available information. It is a powerful tool for combining evidence and changing prior knowledge in the presence of new evidence. The Dempster-Shafer theory can be considered as a generalization of the Bayesian theory of subjective probability.

In the following paragraphs, we give a brief introduction to the basic notions of the Dempster-Shafer theory (frequently called theory of belief functions or theory of evidence).

#### Basic Notions

Considering a finite set referred to as *the frame of discernment*  $\Omega$ , a *basic belief assignment (BBA)* is a function  $m: 2^\Omega \rightarrow [0,1]$  so that

$$\sum_{A \subseteq \Omega} m(A) = 1 \quad (1)$$

where  $m(\emptyset) = 0$ , see [16]. The subsets of  $2^\Omega$  which are associated with non-zero values of  $m$  are known as *focal elements* and the union of the focal elements is called *the core*. The value of  $m(A)$  expresses the proportion of all relevant and available evidence that supports the claim that a particular element of  $\Omega$  belongs to the set  $A$  but not to a particular subset of  $A$ . This value pertains only to the set  $A$  and makes no additional claims about any subsets of  $A$ . We denote this value also as a *degree of belief* (or *basic belief mass - BBM*).

Shafer further defined the concepts of *belief* and *plausibility* [16] as two measures over the subsets of  $\Omega$  as follows:

$$Bel(A) = \sum_{B \subseteq A} m(B), \quad (2)$$

$$Pl(A) = \sum_{B \cap A \neq \emptyset} m(B). \quad (3)$$

A BBA can also be viewed as determining a set of probability distributions  $P$  over  $\Omega$  so that  $Bel(A) \leq P(A) \leq Pl(A)$ . It can be easily seen that these two measures are related to each other as  $Pl(A) = 1 - Bel(\neg A)$ . Moreover, both are equivalent to  $m$ . Thus, one needs to know only one of the three functions  $m$ ,  $Bel$ , or  $Pl$  to derive the other two. Hence, we can speak about belief function using corresponding BBAs in fact.

*Dempster's rule of combination* can be used for pooling evidence represented by two belief functions  $Bel_1$  and  $Bel_2$  over the same frame of discernment coming from independent sources of information. The Dempster's rule of combination for combining two belief functions  $Bel_1$  and  $Bel_2$  defined by (equivalent to) BBAs  $m_1$  and  $m_2$  is defined as follows (the symbol  $\oplus$  is used to denote this operation):

$$(m_1 \oplus m_2)(A) = \frac{1}{1-k} \sum_{B \cap C = A} m_1(B) \cdot m_2(C), \quad (4)$$

where

$$k = \sum_{B \cap C = \emptyset} m_1(B) \cdot m_2(C). \quad (5)$$

Here  $k$  is frequently considered to be a *conflict measure* between two belief functions  $m_1$  and  $m_2$  or a measure of conflict between  $m_1$  and  $m_2$  [16]. Unfortunately, this interpretation of  $k$  is not correct, as it includes also internal conflict of individual belief functions  $m_1$  and  $m_2$  [5, 6]. Dempster's rule is not defined when  $k = 1$ , i.e. when cores of  $m_1$  and  $m_2$  are disjoint. This rule is commutative and associative; as the rule serves for the cumulation of beliefs, it is not idempotent.

#### Belief Function Correction

When receiving a piece of information represented by a belief function, some metaknowledge regarding the quality or reliability of the source that provides the information, can be available. In the following paragraphs, we describe briefly some possibilities how to correct the information according to this metaknowledge.

#### Discounting

To handle the lower reliability of information sources, a discounting scheme has been introduced by Shafer [24]. It is expressed by equations:

$${}^\alpha m(A) = \begin{cases} (1-\alpha) \times m(A) & \text{if } A \subset \Omega \\ \alpha + (1-\alpha) \times m(\Omega) & \text{if } A = \Omega \end{cases} \quad (6)$$

where  $\alpha \in [0,1]$  is a discounting factor and  ${}^\alpha m(A)$  denotes the discounted mass of  $m(A)$ . The larger  $\alpha$  is, the more masses are discounted from  $A \subset \Omega$ , while the more mass is assigned to the frame of discernment  $\Omega$ .

## IV. RESULTS AND DISCUSSION

The idea of discounting mechanism is a weakening of a given belief function (BBA). Thus, the principle of the discounting is transferring of parts of basic belief masses (BBMs) of all focal elements which are proper subsets of the frame of discernment to the entire frame. This process is the result of some additional information saying that the source is not entirely reliable. The transfer of BBMS from a source to the framework reflects an increase of the degree of uncertainty regarding the data that the source produces.

#### Use of belief function theory for ground truth estimation

Traditional data fusion processing based on Dempster-Shafer theory consists of obtaining of BBAs due to some mathematical model in the first step. The second step is the discounting of some BBAs which we know about that they are less reliable (6). The final step is the integration of BBAs using a Dempster's rule (4) or using some other suitable combination rule [11, 17, 18, 21]. As it was described above discounting process is used when we have meta-information about the reliability of some contextual sources of information (BBA) and it is necessary to have some approach how to express the value of discounting factor [1, 2].

In the most cases, the discount rate is adjusted manually, but some authors have suggested several methods how to obtain them automatically. In [18], Smets calculates the discount factor by minimizing the error function. This method focuses on the classification of data and requires a set of labeled data. In [12], Martin et al. establish the discount rate evaluation method that is based only on the values of BBA themselves. Similar approach which is the basis of our work is presented in [10].

Defining what the majority opinion means within the Dempster-Shafer theory is not easy. Murphy [13] for example suggested using average BBAs and argued that the average properties are better suited for the fusion of contradictory evidence:

$$m_{mean} = \frac{1}{M} \sum_{i=1}^M m_i \quad (7)$$

This opinion is valid considering the fact that if subset  $s_1$  from  $S$  corresponds to the cluster of concordant BBAs and if this subset contains more BBAs than any other cluster, then  $m_{mean}$  will probably be closer to BBAs forming the  $s_1$ . Hence  $m_{mean}$  can be used as an estimate of the majority opinion [13]. We therefore propose to review the first set of discount factors by the following way:

$$\alpha_i^0 = d_{BPA}(m_i, m_{mean}) \quad (8)$$

where  $d_{BPA}$  is defined subsequently [12]:

$$d_{BPA}(m_1, m_2) = \sqrt{1/2(\vec{m}_1 - \vec{m}_2)^t D(\vec{m}_1 - \vec{m}_2)} \quad (9)$$

Here, the  $\vec{m}$  is BBA expressed in the form of vector and  $D$  is the matrix which has dimensions  $2^N \times 2^N$  with elements  $D(A,B) = |A \cap B|/|A \cup B|$ .

Equation (10) gives low values of discounting factor for BBAs near to the mean (they are in accordance with the opinion of the majority) and a high degree of discounting factor for BBAs that differ considerable from the mean (the ones that are the cause of disagreement).

In this paper, we use an iterative method for calculating of discounting factors. In the first step, discounting factors are calculated for each member of the initial settings using equation (8). Then this iteration process is applied on the BBAs set  $S_1$ . New values of discounting factors are obtained. This iteration is repeated and the value of discounting factors increases but more and more slowly. To determine the optimal set of discount factors among those computed at each iteration step a posteriori analysis is employed.

We investigate the conjunctive combinations obtained at each step and compare them with categorical BBAs by distance  $d_{BPA}$ . Iteration that gives minimum distance is optimal number of iteration  $i_{opt}$ .

Relative values of discount factors in single steps affect the result of the result of information fusion process as much as the absolute value. In other words, it is not sufficient to have a high degree of value on unreliable sources, it is also necessary that the measure of the difference between reliable and unreliable sources be large enough. Therefore, we perform the optimum setting of values  $\alpha_i$  using iteration. We calculate a discounting factor of the initial set of BBAs and then recalculate new values of BBAs of this set. This process is repeated as described in the previous paragraph. Consecutive values of discount factors are calculated by these iterations process and are further analyzed to determine the best setting according to the predefined criteria which is minimum distance.

An iterative procedure involves the gradual discounting the original BBAs. The term  $m^{\alpha^0, \alpha^1}$  indicates BBA discounted value of  $\alpha^1$ . Successive values of discounting factors  $\{\alpha^0, \dots, \alpha^K\}$  can be summarized:

$$\left. \begin{aligned} \alpha^K &= 1 - \prod_{i=0}^K (1 - \alpha^i) \\ \alpha^K &= \beta^{K-1} (1 - \alpha^K) + \alpha^K \end{aligned} \right\} \quad (10)$$

Stop condition is distance  $d_{BPA}$ . Iteration that gives minimum distance is optimal number of iteration  $i_{opt}$ . Important here is that we can also find a source that differs mostly from the average value. It may be omitted from the calculations and it may be explored independently. The advantage of this described approach is that it does not need any meta-information about the reliability of sources.

TABLE 1. THE BBA SET AND THE RESULTS OF AGGREGATION

|       | {a}    | {b}    | {c}    | {a, b} | {a, c} | {b, c} | {Ω}    |
|-------|--------|--------|--------|--------|--------|--------|--------|
| $m_1$ | 0.5    | 0.2    | 0.1    | 0.1    | 0.05   | 0.025  | 0.025  |
| $m_2$ | 0.52   | 0.12   | 0.08   | 0.05   | 0.1    | 0.06   | 0.07   |
| $m_3$ | 0.6    | 0.08   | 0.12   | 0.025  | 0.1    | 0.025  | 0.05   |
| $m_4$ | 0.2    | 0.1    | 0.6    | 0.05   | 0.025  | 0.025  | 0      |
| $m_5$ | 0.48   | 0.15   | 0.09   | 0.13   | 0.04   | 0.02   | 0.09   |
| $m_6$ | 0.45   | 0.21   | 0.11   | 0.09   | 0.06   | 0.05   | 0.03   |
| $m^*$ | 0.3989 | 0.1022 | 0.2762 | 0.0598 | 0.0785 | 0.0621 | 0.0223 |
| $m$   | 0.5112 | 0.1899 | 0.1169 | 0.0895 | 0.049  | 0.0234 | 0.0201 |

The responses of various sources (observers) are represented by the values of belief functions in Table 1. The six different sources are modeled ( $m_1 - m_6$ ). Ground truth has the same values as the values  $m_1(\cdot)$ . The value of  $m^*(\cdot)$  is calculated using equation (4). The value of  $m(\cdot)$  in the last but one row of the table is calculated according to the process outlined in the previous section. Source 4 ( $m_4$ ) is modeled as adversarial, because its reaction is opposite to the ground truth. The discount factor calculated for this source reaches the highest values. The table shows that discounting process overrides the impact of this source and as a result the result of the integration of information sources will be close to grand truth ( $m^*$ ).

## V. CONCLUSION

This article examines the problem of multiple observers which provide answers that are not entirely accurate. The problem concerns the use of model that is based on belief function theory and no additional information about the reliability of observers are known. Our approach provides an estimate of the ground truth and predicts the response of each observer of the new instance. Experiments show that the proposed method outperforms several core values and leads to a performance close to the model trained with ground truth. There are many opportunities for further research. One possible direction is to extend our model with more cores learning. The aim is to choose an algorithm or a composite different covariance functions instead of fixing the combination in advance. Consequently, the algorithm may be difficult to learn fits observer selecting multiple cores in data-dependent manner. In addition, it would be very useful to design efficient sampling methods for selection that instance and the response should be taught more. Our aim is to test further the described algorithm on real data and further to verify the model described in this paper.

## REFERENCES

- [1] Beranek, L., Nydl, V., The Use of Belief Functions for the Detection of Internet Auction Fraud. In: *Proceedings of the 31st International Conference Mathematical Methods in Economics* 2013. Jihlava: College of Polytechnics Jihlava, 2013, pp. 31 - 36.
- [2] Beranek, L., Knizek, J., The Usage of Contextual Discounting and Opposition in Determining the Trustfulness of Users in Online Auctions. *Journal of Theoretical and Applied Electronic Commerce Research*, Vol. 7, No. 1, 2012, pp. 34-50.
- [3] Chen, S., Zhang, J., Chen, G., Zhang, C., What if the irresponsible teachers are dominating? In Proc. 24<sup>th</sup> AAAI 2010, 2010.
- [4] Crammer, K., Kearns, M., Wortman, J., Learning from multiple sources. *JMLR*, Vol. 9, No. 4, 2008, pp. 1757-1774.
- [5] Daniel, M., Several Notes on Belief Combination, In *Proceedings of the Theory of Belief Functions Workshop*. Brest: ENSIETA, 2010. pp. 1–5.
- [6] Daniel, M., Conflicts within and between Belief Functions. In Proceeding from IPMU 2010 – *Computational Intelligence for Knowledge-Based Systems Design*, Berlin, 2010, s. 696–705, Lecture Notes in Artificial Intelligence. 6178.
- [7] Dawid, A., Skene, A., Maximum likelihood estimation of observer error-rates using the EM algorithm. *Applied Statistics*, Vol. 28, No 1, 1979, pp. 20-28.
- [8] Han, X., Huang, X., Eckert, C., Learning from Multiple Observers with Unknown Expertise, In *Proceedings of 17th Pacific-Asia Conference on Knowledge Discovery and Data Mining*, Gold Coast, Australia, 2013, pp. 233–241
- [9] Hui, S., Walter, S., Estimating the error rates of diagnostic tests. *Biometrics*, Vol. 5, No , 1980, pp. 167171.
- [10] Klein, J., Colot, O., Automatic discounting rate computation using a dissent criterion, In *Proceedings of the Theory of Belief Functions Workshop*. Brest: ENSIETA, 2010. pp. 151–156.
- [11] Lysek, J., Stastny, J., Motycka, A., Object Recognition by Means of Evolved Detector and Classifier Program. In MENDEL 2012, 18th International Conference on Soft Computing. Brno University of Technology, 2012, p. 82–87.
- [12] Martin, A., Jouselme, A.L., Osswald, C., Conflict measure for the discounting operation on belief functions. In *IEEE Int. Conf. on Information Fusion FUSION* 2008, Madrid, 2008, pp. 1-8.
- [13] Murphy, C., Combining belief functions with evidence conflicts. *Decision Support Systems*, Vol. 29, No. 4, 2000, pp. 1-9.
- [14] Raykar, V., Yu, S., Zhao, L., Jerebko, A., Florin, C., Valadez, G., Bogoni, L., Moy, L., Supervised learning from multiple experts: Whom to trust when everyone lies a bit. In *Proc. 26th ICML 2009*, 2009, pp. 889-896. ACM (2009)
- [15] Raykar, V., Yu, S., Zhao, L., Valadez, G., Florin, C., Bogoni, L., Moy, L., Learning from crowds. *JMLR*, Vol. 11, No. 2, 2010, pp. 1297-1322.
- [16] Shafer, G., A mathematical theory of evidence, *Princeton University Press*, Princeton, 1975.
- [17] Schubert, J., Conflict management in Dempster- theory by sequential discounting using the degree of falsity. In *Int. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems IPMU 2008*, Madrid, 2008, pp. 298-305.
- [18] Smets, P., Analyzing the combination of conflicting belief functions. *Information Fusion*, Vol. 8, No. 3, 2006, pp. 387-412.
- [19] Smyth, P., Fayyad, U., Burl, M., Perona, P., Baldi, P., Inferring ground truth from subjective labelling of venus images. In *Proc. 9th NIPS 1995*, 1995, pp. 1085-1092.
- [20] Spiegelhalter, D., Stovin, P., An analysis of repeated biopsies following cardiac transplantation. *Statistics in Medicine*, Vol. 2, No. 1, 1983, pp. 33-40.
- [21] Stencl, M., Stastny, J., Neural network learning algorithms comparison on numerical prediction of real data. In *MENDEL 2010, 16th International Conference on Soft Computing*. Brno University of Technology, 2010, p. 280-285.
- [22] Tubaishat, M., Madria, S., Sensor networks: an overview. *Potentials IEEE*, Vol. 22, No. 2, 2003, pp. 20-23.
- [23] Whitehill, J., Ruvolo, P., Wu, T., Bergsma, J., Movellan, J., Whose vote should count more: Optimal integration of labels from labelers of unknown expertise. In *Proc. 23rd NIPS 2009*, vol. 22, pp. 2035-2043.
- [24] Wu, O., Hu,W., Gao, J.: Learning to rank under multiple annotators. In: *22nd IJCAI*, 2011, pp. 1161–1168.
- [25] Yan, Y., Rosales, R., Fung, G., Dy, J.: Active learning from crowds. In: *Proc. 28th ICML*, 2011, pp. 1161–1168.
- [26] Yan, Y., Rosales, R., Fung, G., Schmidt, M., Hermosillo, G., Bogoni, L., Moy, L., Dy, J., Malvern, P., Modeling annotator expertise: Learning when everybody knows a bit of something. *Journal of Machine Learning Research*, Vol. 9, 2010, pp. 932-939.

# Fuzzy Controller of Pathological Conditions Diagnosis based on Analysis of Cytological Images

Lesia Dubchak<sup>1</sup>, Serhiy Verbovyy<sup>2</sup>, Olena Verbova<sup>2</sup>, Nadiia Vasylkiv<sup>3</sup>

1,2. Department of Computer Engineering, Ternopil National Economic University, UKRAINE, Ternopil, 8 Chekhova str.,  
Email: [dlo@tneu.edu.ua](mailto:dlo@tneu.edu.ua)<sup>1</sup>, [vso@tneu.edu.ua](mailto:vso@tneu.edu.ua)<sup>2</sup>

3. Department of Information and Computing Systems and Control, Ternopil National Economic University, UKRAINE, Ternopil, 11 Lvivska str., Email: [nvs@tneu.edu.ua](mailto:nvs@tneu.edu.ua)

**Abstract:** In this article authors proposed the fuzzy controller for the correct diagnosis of the breast pathological states. This mean can be used in medical practice by the cytologist as an additional way of diagnosis confirming that is the vital sometimes.

**Keywords:** fuzzy controller, breast cancer, cytological images, pathological conditions.

## I. INTRODUCTION

Breast cancer in Ukraine, as well as in most countries of the world, is one of the most widespread tumor processes in women and ranks first in the structure of morbidity and mortality from malignant neoplasms (25% of all cases of cancer).

The cytological research allows to identify pathological changes of cells at early stages of development, since the main object of the study are small cellular structures, namely: nucleus, cytoplasm, mitochondria, as well as determination of nuclear-cytoplasmic ratio, which is a very important indicator. The analysis of the structure of cells is carried out using powerful microscopes with high resolution and high quality optical components.

The cytological analysis is carried out for the purpose of screening of oncological diseases, for the form verification of the pathological process (proliferative, atrophic, atypical, inflammatory, autoimmune), the effectiveness of the performed treatment by identifying the regression of the disease, to control the dynamics of changes in the cells.

The prevalence of cytological images processing software allows the process of establishing the exact diagnosis to be the fastest and most accurate for an expert. To determine the exact diagnosis based on qualitative signs of images, excluding the subjectivity of judgments of the expert, it is worth using a fuzzy logic system.

## II. FUZZY SYSTEM

In oncology, fuzzy logic, usually in combination with neural networks, is used mainly for processing the images themselves or some markers of the occurrence of pathological states. For example, Alberto d'Onofrio in his work on the use of fuzzy logic in medicine within the study of the impact of chemotherapy indicates that although the fuzzy system has its disadvantages, but its outcome is biologically more realistic than other approaches.

A group of scientists of the State University of Louisiana also investigated the use of fuzzy logic in order to exclude the

subjectivity of decision making of radiologists when applying the terminology and vocabulary of breast cancer diagnostics.

Another area of using the fuzzy logic is the processing of medical imaging of tissues of various organs. The main advantage of this approach is the rapid training of the system and the high probability of the results [1].

Fuzzy systems can operate incoming information that is unclear, for example, verbally, by expert-physician, perform fuzzy formalizations of evaluation and comparison criteria, carry out qualitative evaluations of input data and output results due to their degree of probability and distribution, carry out rapid simulation of complex dynamic systems and their comparative analysis with a given degree of accuracy. Therefore, this approach will allow to build a diagnostic system for the pathological states of the mammary gland, which will work in real time and is capable of quick adjustment.

The membership function determines the subjective confidence level of the expert in that the given specific value of the base scale corresponds to the value of the fuzzy set. The setting of production rules is based on the definition of such fuzzy rules, so that the control module, constructed on their basis, generates certain output signals when receiving input signals. So that is why, it is necessary to divide the space of incoming and outgoing signals into sets and set the appropriate membership functions for them. Record fuzzy rules based on the experimental sample, create a table for writing the production rule base and the truth table of the rules (presence or absence of signs). Set all the rules of the degree of truth and generate the rules, form the basis of fuzzy rules [2].

Fuzzy modeling in the Matlab setting is based on the application of the Fuzzy Logic Toolbox extension package, which presents a large number of fuzzy logic functions and fuzzy output.

The fuzzy system of processing cytological images for the diagnosis of pathological states of the mammary gland is based on the mechanism of the Mamdani fuzzy conclusion, in which the knowledge data base consists of the rules of the form "if-then". In this case, the input variables are the signs of the pathological states that are present in the image [3-5]. For ease of construction, the following designations are used:

- c1 - flattened apocrine epithelium;
- c2 - formation of papillary structures;
- c3 - the presence of secretory activity in cells;
- c4 - round hyperchromic nuclei, located centrally;

- c5 - a small number of hyperchromic monomorphic cells;  
 c6 - the cells are layers;  
 c7 - in the background there are many phagocytes and histiocytes;  
 c8 - the presence of secretions around the cellular space;  
 c9 - formation of cellular complexes;  
 c10 - formation of papillary complexes with dense placement of cells in multilayered layers;  
 c11 - large cell sizes;  
 c12 - large sizes of nuclei with intensively expressed chromatin;  
 c13 - intensely expressed nuclei;  
 c14 - narrow rim of intensively painted cytoplasm;  
 c15 - round hyperchromic nuclei;  
 c16 - fibroblasts.

It is proposed that each of the signs is given only by two fuzzy states "present" or "missing" in the image.

For the diagnosis of nonproliferative, proliferative mastopathy and fibroadenomas, there are no cytological images of mutually exclusive features. However, there are some signs that may be present at the same time or must be present.

In particular, to confirm the diagnosis of nonproliferative mastopathy in a cytological image, the diagnostician must always observe the signs c1, c3, c4 and c5. In addition, the signs are preferably at the same time c2, c5, c6 or c1, c3, or c1, c3, c8. On that basis we can conclude that in order to confirm the diagnosis of nonproliferative mastopathy, it is necessary to work out a database of 16 rules.

The proliferative mastopathy and fibroadenoma can only be confirmed by the physician if there are simultaneously signs of c9, c10, c11 and c12 (in the case of mastopathy) and c1, c2, c11, c13, c14, c15, c16 (in the case of fibroadenoma). In other words, in general, the fuzzy system is working on the basis of 18 rules of the type "if, then":

1. If (c1 is present) and (c3 is present) and (c4 is present) and (c5 is present) then (diagnosis-cytology is nonproliferative mastopathy).
2. If (c1 is present) and (c2 is present) and (c3 is present) and (c4 is present) and (c5 is present) then (diagnosis-cytology is nonproliferative mastopathy).
3. If (c1 is present) and (c3 is present) and (c4 is present) and (c5 is present) and (c6 is present) then (diagnosis-cytology is nonproliferative mastopathy).
4. If (c1 is present) and (c3 is present) and (c4 is present) and (c5 is present) and (c7 is present) then (diagnosis-cytology is nonproliferative mastopathy).
5. If (c1 is present) and (c3 is present) and (c4 is present) and (c5 is present) and (c8 is present) then (diagnosis-cytology is nonproliferative mastopathy).
6. If (c1 is present) and (c2 is present) and (c3 is present) and (c4 is present) and (c5 is present) and (c6 is present) then (diagnosis-cytology is nonproliferative mastopathy).
7. If (c1 is present) and (c2 is present) and (c3 is present) and (c4 is present) and (c5 is present) and (c7 is present) then (diagnosis-cytology is nonproliferative mastopathy).
8. If (c1 is present) and (c2 is present) and (c3 is present) and (c4 is present) and (c5 is present) and (c8 is present) then (diagnosis-cytology is nonproliferative mastopathy).

9. If (c1 is present) and (c3 is present) and (c4 is present) and (c5 is present) and (c6 is present) and (c7 is present) then (diagnosis-cytology is nonproliferative mastopathy).
10. If (c1 is present) and (c3 is present) and (c4 is present) and (c5 is present) and (c6 is present) and (c8 is present) then (diagnosis-cytology is nonproliferative mastopathy).
11. If (c1 is present) and (c3 is present) and (c4 is present) and (c5 is present) and (c7 is present) and (c8 is present) then (diagnosis-cytology is nonproliferative mastopathy).
12. If (c1 is present) and (c2 is present) and (c3 is present) and (c4 is present) and (c5 is present) and (c6 is present) and (c7 is present) then (diagnosis-cytology is nonproliferative mastopathy).
13. If (c1 is present) and (c2 is present) and (c3 is present) and (c4 is present) and (c5 is present) and (c6 is present) and (c8 is present) then (diagnosis-cytology is nonproliferative mastopathy).
14. If (c1 is present) and (c2 is present) and (c3 is present) and (c4 is present) and (c5 is present) and (c7 is present) and (c8 is present) then (diagnosis-cytology is nonproliferative mastopathy).
15. If (c1 is present) and (c3 is present) and (c4 is present) and (c5 is present) and (c6 is present) and (c7 is present) and (c8 is present) then (diagnosis-cytology is nonproliferative mastopathy).
16. If (c1 is present) and (c2 is present) and (c3 is present) and (c4 is present) and (c5 is present) and (c6 is present) and (c7 is present) and (c8 is present) then (diagnosis-cytology is nonproliferative mastopathy).
17. If (c9 is present) and (c10 is present) and (c11 is present) and (c12 is present) then (diagnosis-cytology is proliferative mastopathy).
18. If (c1 is present) and (c2 is present) and (c11 is present) and (c13 is present) and (c14 is present) and (c15 is present) and (c16 is present) then (diagnosis-cytology is fibroadenoma).

FuzzyLogicToolbox Matlab setting is used to build a fuzzy system.

The input variables of this fuzzy system are the signs c1-c16 described above. The output of the proposed system (diagnosis-cytology) is the diagnosis of nonproliferative, proliferative mastopathy and fibroadenoma. The general view of the fuzzy system for the diagnosis of a cytological image is given in Fig. 1.

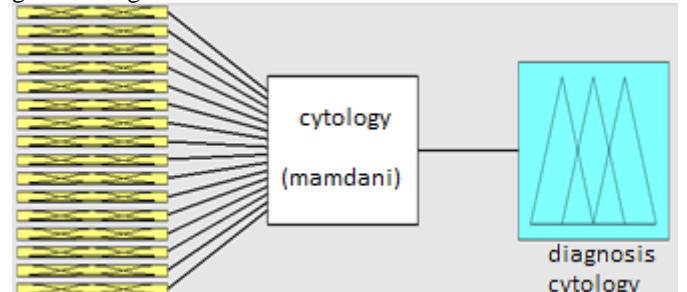


Fig. 1. A general view of the developed fuzzy system.

The membership functions of the input variables, that is the signs c1-c16, are given by a bell-shaped form that best reflects the two sets of values of each of them, namely, the present or the missing attribute on the image (Fig. 2) [17].

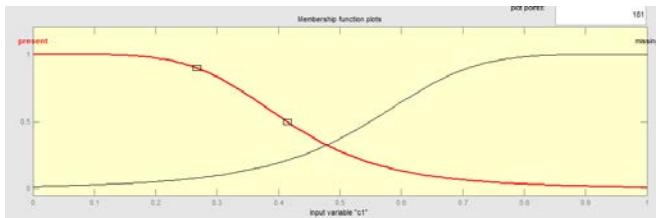


Fig. 2. The membership functions of input variables for example c1 signs.

To set the membership functions, a triangular shape is used (Fig. 3).

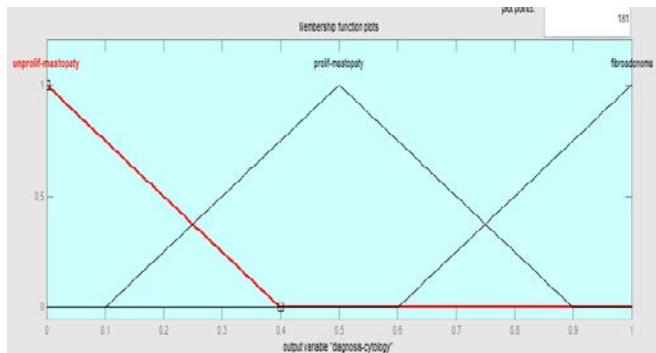


Fig. 3. The membership functions of target variable diagnosis-cytology.

The correctness of the work of the developed fuzzy system follows from the analysis of the fuzzy conclusion obtained when the given rule base is working (Fig. 4).

On the basis of the developed rules of the fuzzy conclusion a system of diagnostics of pathological states of the mammary gland was constructed.

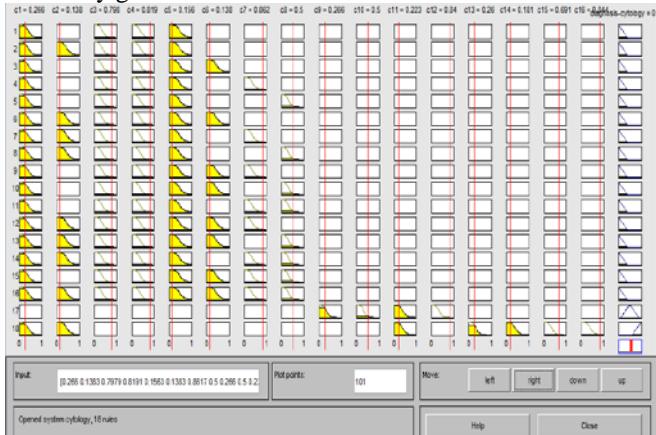


Fig. 4. The result of the fuzzy diagnostic system of pathological states of the mammary gland.

### III. FUZZY CONTROLLER

The basic fuzzy controller consists of four main components:

- Fuzzyfication unit (just changes the inputs, so they can be interpreted and compared with the rules from the base knowledge);
- Knowledge base (base rules and database that holds knowledge, in the form of a set of rules, on how to better manage the system);
- decision-making unit (logical conclusion mechanism, which evaluates which rule is now relevant, and then decides

what should be submitted to the input);

• unit of defuzzification (transmits the conclusions made by using the mechanism of logical conclusion, to the inputs). The model of the fuzzy controller of access to the system of assessment of a general educational institution can be done by using Simulink. Simulink is an interactive tool for modeling, simulating and analyzing dynamic systems, including discrete, continuous and hybrid, nonlinear and discontinuous systems.

The general scheme of the fuzzy controller is shown in Fig. 5.

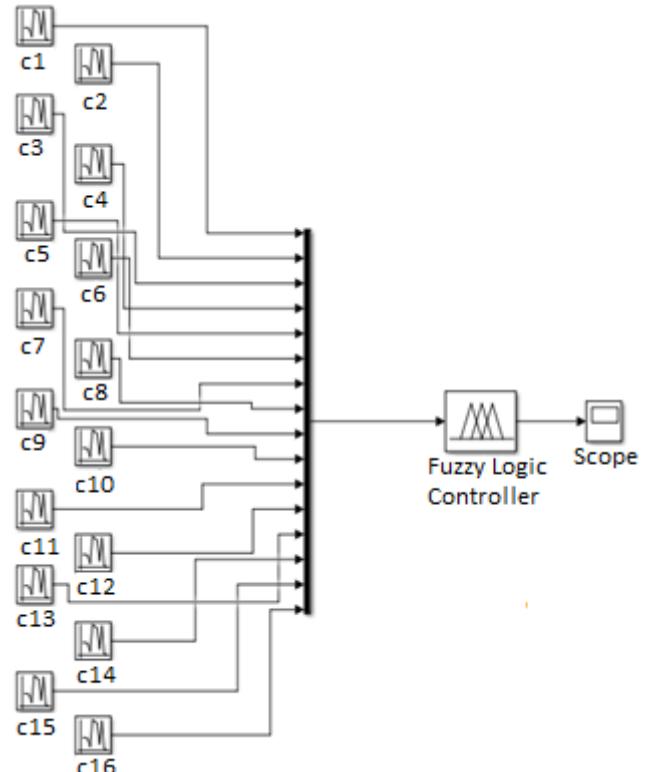


Fig. 5. General scheme of the fuzzy controller.

Input variables are set like Random Number.

General scheme of input and output values are shown in Fig. 6.

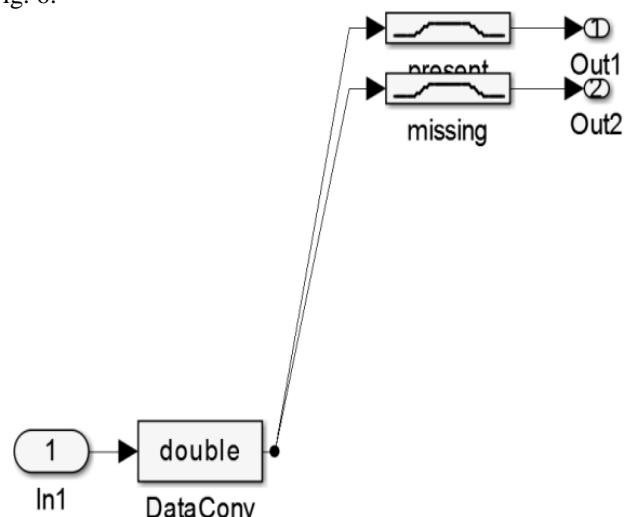


Fig. 6. General scheme of input values.

Fig. 7 depicts the scheme for processing incoming fuzzy

values by the rule of the type “if-then”. Simulink handles rules from the knowledge base, taking into account the rating displayed by the constant Weight.

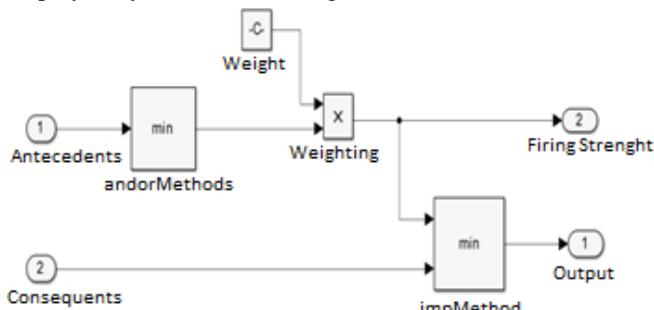


Fig. 7. Scheme for handling incoming fuzzy values using the rule “if-then”.

To make a conclusion on the Mamdani mechanism, the fuzzy controller carries out defuzzification. Defuzzification scheme is shown in Fig. 8.

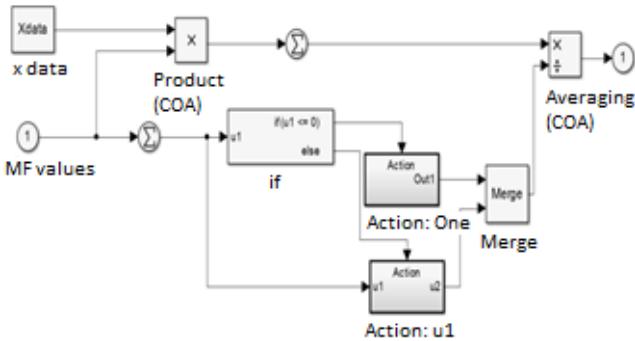


Fig. 8. Defuzzification scheme of the fuzzy conclusion.

In order to analyze the work of the fuzzy controller, it is necessary to use data from the Scope units, which are shown in Fig. 9.

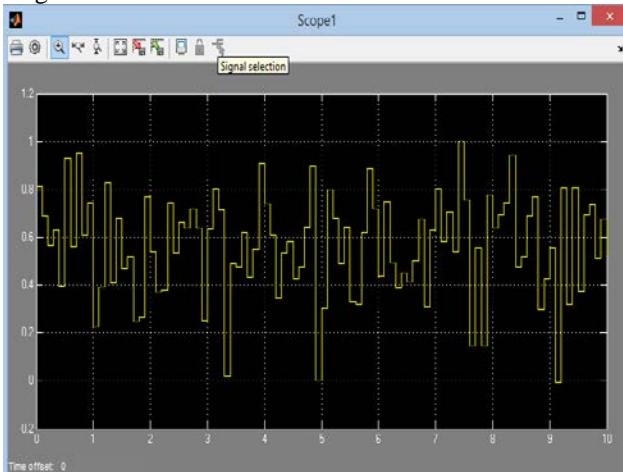


Fig. 9. Evenly distributed task of random values of the input variable c1.

The result of the model at given input values of the presence of breast pathological states signs is shown in Fig. 10.

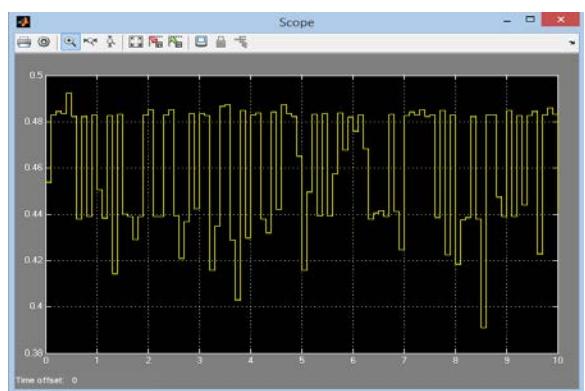


Fig. 10. Results of the developed fuzzy controller.

According to the data presented in Fig. 9 and 10, it can be assumed that the developed fuzzy controller is working correctly and can be used with aim of the correct diagnosis of the breast pathological states. Moreover, this controller can be coded in VHDL with help of Simulink and realized on FPGA.

#### IV. CONCLUSION

The given fuzzy system solves the basic problem of diagnosing the pathological states of the mammary gland, namely, the subjectivity of the judgments of the diagnostician in the analysis of cytological images. In addition, such a system can be used for doubtful diagnoses as confirmation of expert opinion.

The proposed fuzzy controller is a prototype of the hardware tool, which can be used in telemedicine system.

#### REFERENCES

- [1] Abadeh M.S., Habibi J., Lucas C. “Intrusion Detection Using a Fuzzy Genetics-Based Learning Algorithm”, *Journal of Network and Computer Applications*, 2007, №30, P.414-428.
- [2] Ross T.J. “Fuzzy Logic with Engineering Applications”, *McGraw-Hill Inc. (USA)*, 1995, 600 p.
- [3] Oleh Berezsky, Oleh Pitsun, Serhiy Verbovyy, Tamara Datsko, Andriy Bodnar “Computer diagnostic tools based on biomedical image analysis”, *Proceedings of the 14th International Conference The Experience of Designing and Application of CAD Systems in Microelectronics, CADSM 2017*, 21-25 February, Polyana-Svalyava, pp. 388-391.
- [4] Lesia Dubchak, Serhiy Verbovyy, Kateryna Berezska, Tamara Datsko “Fuzzy knowledge base for diagnosing breast cancer pathological processes”, *Proceedings of the XII International Scientific and technical conference Computer science & information technologies (CSIT 2017)*, 5-8 September 2017, Lviv, Ukraine - pp. 36-39.
- [5] L.O.Dubchak, S.O.Verbovyy, O.M.Berez'kyy, T.V.Datsko, N.YA. Savka “Nechitka sistema diahnostuvannya patolohichnykh staniv molochnoyi zalozy na osnovi analizu tsytolohichnykh zobrazen”, *Visnyk Khmel'nyts'koho natsional'noho universytetu*, 2017. - №5 (253). - S. 203-211. (in Ukrainian).

# The Associative Rules Constructing on the Example of Patient's Physical Characteristics

Nataliya Shakhovska, Iryna Zhelizniak

Department of Artificial Intelligence, Lviv Polytechnic National University, UKRAINE, Lviv, 12 S.Bandera str.,  
email: [nataliya.b.shakhovska@lpnu.ua](mailto:nataliya.b.shakhovska@lpnu.ua)

**Abstract:** The method of the construction of associative rules are described. Associative rules for assaying the patient have been constructed. The set of transactions that are available for medical analysis of a patient is considered. It has been found that the correct assessment of the utility of an associative rule affects the volume and speed of access to information. A unique identifier for the patient set of patient analyzes has been entered. Additional numerical attributes of the investigated objects are indicated. Transactions that contain additional attributes and operations are not only available, but also compared. The distinction between associative rules and sequential analysis is given.

**Keywords:** associative rules, data mining, support, patient's physical characteristics, sequential analysis

## I. INTRODUCTION

In medical and biological research, as well as in practical medicine, the range of tasks to be solved is so wide that it is possible to use any of the methodologies of Data Mining. An example can be the construction of a diagnostic system or the study of the effectiveness of surgical intervention.

One of the most advanced areas of medicine is bioinformatics. The object of bioinformatics research is huge amounts of information about DNA sequences and the primary structure of proteins that arose as a result of studying the structure of genomes of microorganisms, mammals and humans. Abstracted from the specific content of this information, it can be regarded as a set of genetic texts, consisting of extended character sequences. Detection of structural laws in such sequences is a number of tasks, effectively solved by means of Data Mining, for example, by means of sequencing and associative analysis [1, 2].

The purpose of the study is to identify the most important rules for constructing associative rules. Determination of the patterns of constructing associative rules and the division of physical indicators at different levels of the hierarchy..

## II. OBJECTS AND METHODS OF RESEARCH

One of the most common data analysis tasks is to identify sets of objects that are often encountered in a large set of objects. We describe this problem in a generalized form. To do this, we denote the objects that make up the study sets (itemsets), as follows [2, 3]:

$$I = \{i_1, i_2, \dots, i_j, \dots, i_n\}, \quad (1)$$

where  $i_i$  - objects included in the studied sets;  $n$  - total number of objects.

In the field of medicine, such objects, for example, are indicators and analyzes of the patient (Table 1).

TABLE 1. OBJECTS INCLUDED IN THE STUDY SET

| <b>Id</b> | <b>Indicator</b>                 | <b>Value</b>   |
|-----------|----------------------------------|----------------|
| 0         | Blood pressure                   | 120/80 mm. Hg. |
| 1         | Venous pressure                  | 70 mm. H2O     |
| 2         | Capillary pressure               | 70 mm. Hg.     |
| 3         | Pulse                            | 85 beats/min   |
| 4         | Temperature                      | 36,6 C         |
| 5         | Level of hemoglobin in the blood | 145 Hb         |
| 6         | pH                               | 7.35           |

In this way they correspond to the following set of objects:  $I = \{\text{arterial pressure, venous pressure, capillary pressure, pulse, temperature, hemoglobin level in blood, pH}\}$ .

Sets of objects from the  $I$  set, stored in a database and subject to analysis, are called transactions. We describe the transaction as a subset of the set  $I$ :

$$T = \{i_j | i_j \in I\}. \quad (2)$$

Such transactions in the hospital are in accordance with the delivery of medical examinations of the patient and stored in the database in the form of a medical card. They list the tests that the patient passed for a history and diagnosis.

The set of transactions, the information about which is available for analysis, will be described by the following set:

$$D = \{T_1, T_2, \dots, T_r, \dots, T_m\}, \quad (3)$$

where  $m$  - the number of transactions available for analysis.

## III. RESEARCH RESULTS

To use Data Mining methods, the set  $D$  can be represented as a table (Table 2).

The set of transactions, which includes  $j_i$  objects, is indicated as follows [3]:

$$D = \{T_r | i_j \in T_r; j = 1..n; r = 1..m\} \subseteq D \quad (4)$$

In this example, the set of transactions containing the Object Temperature is the following:

TABLE 2. A SET OF INVESTIGATED OBJECTS

| Transaction number | Indicator number | Indicator       | Value                   |
|--------------------|------------------|-----------------|-------------------------|
| 0                  | 0                | Blood pressure  | 110/75 mm. Hg.          |
| 0                  | 3                | Pulse           | 110 beats/min           |
| 0                  | 1                | Venous pressure | 58 mm. H <sub>2</sub> O |
| 1                  | 4                | Temperature     | 37.4 C                  |
| 1                  | 5                | pH              | 7.46                    |
| 2                  | 1                | Venous pressure | 72 mm. H <sub>2</sub> O |
| 2                  | 6                | pH              | 7.81                    |
| 2                  | 4                | Temperature     | 37.2 C                  |

In this example, the set of transactions containing the Object Temperature is the following set:

$$D_{\text{temperature}} = \{\{\text{Temperature}, \text{pH}\},$$

$$\{\text{Venous pressure}, \text{pH}, \text{Temperature}\}$$

Some arbitrary set of objects (itemset) is denoted as follows:

$$F = \{i_j | i_j \in I; j = 1..n\}. \quad (5)$$

The set of transactions that includes the set F is denoted as follows:

$$D_F = \{T_r | F \subseteq T_r; r = 1..m\} \subseteq D. \quad (6)$$

The ratio of the number of transactions, which includes the set F, to the total number of transactions is called support of the set F and denoted by Supp(F):

$$\text{Supp}(F) = |D_F|/D. \quad (7)$$

For example, for a set {pH, temperature} the subtraction will be equal to 2/3, because this set is included in two transactions (numbers 1 and 2) of the three possible.

When searching, an analyst can specify the minimum value of maintaining interesting sets - Suppmin. A set is called large if its value exceeds the minimum support value specified by the user:

$$\text{Supp}(F) > \text{Supp}_{\min}. \quad (8)$$

So, when searching for associative rules you need to find the set of all frequent sets:

$$L = \{F | \text{Supp}(F) > \text{Supp}_{\min}\}. \quad (9)$$

In this case, the sets with Suppmin = 2/3 are the following:

$$\{\text{Venous pressure}\} \text{Suppmin} = 2/3;$$

$$\{\text{Temperature}\} \text{Suppmin} = 2/3;$$

$$\{\text{pH, Temperature}\} \text{Suppmin} = 2/3;$$

In an analysis, the sequence of events is often of interest. When detecting regularities in such sequences, it is possible to predict with some degree the occurrence of events in the future, which allows us to make more correct decisions. A sequence is called an ordered set of objects. To do this, the order must be given to the set [4].

Then the sequence of objects can be described as follows:

$$S = \{\dots, i_p, \dots, i_q\}, \text{where } p < q. \quad (10)$$

For example, in the case of analyzes such a sequence of objects may be the date of delivery of analyzes. Such a sequence:

$$S = \{\{\text{hemoglobin level, 10.10.2017}\}, \{\text{venous pressure, 09/25/2017}\}, \{\text{pH, 28.09.2017}\}\}$$

can be interpreted as a sequence of delivery of tests by one person at different times (initially measured venous pressure, then measured the pH level, and finally the level of hemoglobin).

There are two types of sequences: with cycles and without cycles. In the first case it is allowed to enter the sequence of the same object at different positions:

$$S = \{\dots, i_p, \dots, i_q, \dots\}, \text{where } p < q, i_q = i_p. \quad (11)$$

It is said that transaction T contains the sequence S, if S ⊆ T and the objects included in S, also belong to the set of T, with preservation of the relation of order. It is supposed that in the set T between objects in the sequence of S there may be other objects.

The maintenance of the sequence S is the ratio of the number of transactions, which includes the sequence of S, to the total number of transactions. The sequence is frequent if its support exceeds the minimum support given by the user:

$$\text{Supp}(S) > \text{Supp}_{\min}. \quad (12)$$

The task of sequential analysis is to search all frequent sequences:

$$L = \{S | \text{Supp}(S) > \text{Supp}_{\min}\}. \quad (13)$$

The main difference between the problems of sequential analysis from the search for associative rules is to establish a relation of order between objects of the set I. This relation can be determined in different ways. In the analysis of the sequence of events occurring in time, the objects of the set I are events, and the order of relationships corresponds to the chronology of their appearance. For example, analyzing sequences of assays in a hospital are sets of analyzes that the patient submits at different times, and the order of reference is the time of the implementation of these analyzes.

D = {{(temperature, blood pressure, capillary pressure)}, {(pH, temperature, pulse)}}, {{(hemoglobin level in blood, temperature), (blood pressure, temperature), (temperature, venous pressure)}}, {{(hemoglobin level in the blood)}}.

Of course, there is a problem of identification of patients. In practice, this is decided by the introduction of medical cards that have a unique identifier (table 3).

TABLE 3. A UNIQUE IDENTIFIER FOR THE SET OF ANALYZES

| Patient ID | Sequence of analyzes delivery   |
|------------|---|
| 0          | (temperature, arterial pressure, capillary pressure), (pH, temperature, pulse)                              |
| 1          | (hemoglobin level in the blood, temperature), (blood pressure, temperature), (temperature, venous pressure) |
| 2          | (hemoglobin level in the blood)   |

The following sequence can be interpreted as follows: the patient with the ID 0 initially passed the temperature, the arterial and capillary pressure, and then passed the pH, temperature and pulse rate with his visit. For example, the support for the  $\{\text{blood pressure, temperature}\}$  sequence is 2/3, since it is found in patients with identifiers 0 and 1.

In many applications, objects of the set I naturally combine into groups that in turn can also be grouped into more general groups, etc. Thus, the hierarchical structure of objects is obtained.

An example of such a hierarchy may be the following categorization of analyzes:

Pressure:

- Arterial;
- Venous;
- Capillary

Physical indicators:

- Temperature

Blood test:

- Hemoglobin level;
- PH

The presence of a hierarchy changes the perception of when an object  $i$  is present in transaction  $T$ . Obviously, support is not a separate object, but the group to which it is included is greater:

$$\text{Supp}(I_q) \geq \text{Supp}(i_j) \quad (14)$$

where  $i_j \in I_q$ .

This is due to the fact that when analyzing groups, not only transactions that include a separate object, but also transactions containing all objects of the analyzed group are counted. For example, if  $\text{Supp} \{\text{blood pressure, temperature}\} = 2/3$ , then support  $\text{Supp} \{\text{pressure, physical parameters}\} = 2/3$ , since the objects of the groups of pressure and physical parameters are included in the transaction with the identifiers 0 and 1.

Using the hierarchy allows you to determine the connection that goes into higher levels of the hierarchy, since the support for the set can increase if the entry of the group, and not its object, is counted. In addition to the search for kits that often occur in transactions, which in turn consist of objects  $F = \{i | i \in I\}$  or groups of the same level of the hierarchy:

$$F = \{I^g | I^g \in I^{g+1}\}. \quad (15)$$

You can also consider mixed sets of objects and groups:

$$F = \{i, I^g | i \in I^g \in I^{g+1}\}. \quad (16)$$

This allows you to extend the analysis and gain additional knowledge.

In the hierarchical structure of objects, you can change the nature of the search by changing the analyzed level. Obviously, the more objects in the set I, the more objects in transactions T and frequent sets. This in turn increases search time and complicates the analysis of results. You can reduce or increase the amount of data using the hierarchical representation of the objects under analysis. Moving up the hierarchy, we summarize the data and reduce their number, and vice versa.

The disadvantage of generalizing objects is the less usefulness of the knowledge gained, since in this case they relate to groups that do not always have useful information. To achieve a compromise between group analysis and analysis of individual objects, they often do the following: first analyze

the groups, and then, depending on the results, investigate the objects that interest the group analyst. In any case, it can be argued that the presence of a hierarchy in objects and its use in the task of finding associative rules allows you to perform a more flexible analysis and gain additional knowledge.

In the considered problem of searching for associative rules, the presence of an object in a transaction was determined only by its presence in it ( $i_j \in T$ ) or the absence ( $i_j \notin T$ ). Often, objects have additional attributes, usually numeric. For example, analyzes in a transaction have attributes: value and duration. In this case, the presence of an object in the set can be determined not only by the fact of its presence, but also the execution of the condition in relation to a certain attribute. For example, in analyzing transactions performed by patients, they are interested not only in the value of the analysis, but also in how well this indicator is stable (long-term).

You can add additional objects to explore the sets in order to extend the analysis capabilities by searching for associative rules. In the general case, they may have a nature different from the main objects. For example, in the case of delivery of tests, you can enter the field of delivery frequency or symptoms that precede the delivery of these particular analyzes.

Solving the problem of finding associative rules, as well as any task, is to process the output and obtain the results. Processing of the initial data is performed by a certain Data Mining algorithm.

The results obtained in solving this problem are accepted in the form of associative rules. In this regard, when searching for them, there are two main stages:

1. Finding all large sets of objects;
2. Generation of associative rules from found large sets of objects.

Associative rules are as follows:

If (condition) then (result)

where condition is usually not a logical expression (as in the classification rules), but a set of objects from the set I, with which associated (associated) objects are included in the result of this rule.

For example, associative rule:

If (blood pressure, pH) then (hemoglobin level)

means that if the patient is measured by arterial pressure and pH level, he also measured by hemoglobin level.

As already noted, in associative rules the condition and the result are objects of the set I:

If  $X$  then  $Y$ ,

where  $X \in I, Y \in I, X \cup Y = \varphi$ .

The main advantage of associative rules is their easy perception by a person and a simple interpretation of programming languages. However, they are not always useful. There are three types of rules:

1. Useful rules - contain valid information that was previously unknown but has a logical explanation. Such rules can be used for making decisions that are beneficial;
2. Trivial rules - contain valid and easily understandable information that is already known. Such rules, although they can be explained, but can not bring any benefits, as they reflect or known laws in the studied area, or the

results of past activity. Sometimes such rules can be used to verify the implementation of decisions taken on the basis of preliminary analysis;

3. Unclear rules - contain information that can not be explained. Such rules can be obtained either on the basis of abnormal values, or deeply hidden knowledge. Directly such rules can not be used for decision making, since their lack of clarity can lead to unpredictable results. For better understanding, further analysis is required.

Associative rules are built on the basis of large sets. So, the rules built on the basis of the set  $F$ , are all possible combinations of objects included in it.

For example, for the set {arterial pressure, temperature, pulse} the following associative rules can be constructed:

If (arterial pressure) then (temperature);

If (arterial pressure) then (pulse);

If (arterial pressure) then (temperature);

If (arterial pressure) then (temperature, pulse);

If (temperature, pulse) then (arterial pressure);

And so on.

Thus, the number of associative rules can be very large and bad for human perception. In addition, not all of the built-in rules carry useful information. To assess their usefulness, the following values are entered:

- Support - shows which percentage of transactions supports this rule (we found rules, where Support is upper then 75%).
- Confidence - shows the probability that the presence of a set  $Y$  in the transaction in the set  $X$  implies (we found rules, where Confidence is upper then 0.5).
- Improvement - indicates whether this rule is useful for research.

These estimates are used when generating rules. An analyst when searching for associative rules specifies the minimum values of these variables. As a result, those rules that do not satisfy these conditions are discarded and are not included in the solution of the problem.

If objects have additional attributes that affect the composition of objects in transactions, and therefore in sets, then they should be taken into account in generated rules. In this case, the conditional part of the rules will not only include verification of the existence of an object in a transaction, but also more complex comparing operations: more, less, includes, etc. The resulting part of the rules may also contain statements about the attribute values. For example, if an indicator is considered topical, then the rules may look like this:

If pH.relevance > 10 days then the level of hemoglobin in the blood.relevance < 3 days.

This rule states that the patient did the pH analysis more than 10 days ago, then probably his analysis of hemoglobin in the blood is valid for no more than 3 days.

The main differences between static and dynamic XML documents are:

#### • Availability of validity period

A static XML document does not contain elements that indicate the expiration date of this document. In contrast, a dynamic XML document initially contains at least one element

that indicates the validity period of a particular version of the document.

#### • Persistence of displayed information

Once created, the information of a static XML document remains valid at all times. Conversely, the version of the dynamic XML document is valid only for the period specified in the corresponding elements. As soon as a new version appears, the information contained in the previous version is replaced.

Most of the work on finding associative rules in static XML documents is related to the use of XML-based algorithms based on the Apriori algorithm. However, there are a number of other approaches.

TABLE 4. REPRESENTATION OF STATIC XML DOCUMENT

```
<?xml version="1.0" encoding="UTF-8"?>
<patient>
    <name>Tom Johnson</name>
    <street>Bandery</street>
    <city>Lviv</city>
    <analyse>
        <analyse_date>15/01/2017</analyse_date>
        <results>
            <temperature>38.2</temperature>
            <venouse_pressure>72</venouse_pressure>
                <pulse>110</pulse>
                <pH>7.46</pH>
                <hemoglobin>145</hemoglobin>
            </results>
        </analyse>
    </patient>
```

### III. CONCLUSION

The task of finding associative rules is to identify sets of objects that are commonly encountered in a large number of objects. The task of sequential analysis is to search for frequent sequences. The main difference between the tasks of sequential analysis from the search for associative rules is to establish a relationship of order between objects. The presence of a hierarchy in objects and its use in the task of finding associative rules allows you to perform a more flexible analysis and obtain additional knowledge. The results of the solution of the problem are presented in the form of associative rules, conditional and the final part of which contains sets of objects.

### REFERENCES

- [1] Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems*, 30(1-7), 107-117.
- [2] Negnevitsky, M. (2005). *Artificial intelligence: a guide to intelligent systems*. Pearson Education.
- [3] Jain, V., Benyoucef, L., & Deshmukh, S. G. (2008). A new approach for evaluating agility in supply chains using fuzzy association rules mining. *Engineering Applications of Artificial Intelligence*, 21(3), 367-385.
- [4] Shakhovska, N., Kaminsky, R., Zasoba, E., & Tsiutsiura, M. (2018). ASSOCIATION RULES MINING IN BIG DATA. *International Journal of Computing*, 17(1), 25-32

# The Neo-Fuzzy Autoencoder for Adaptive Deep Neural Systems and its Learning

Yevgeniy Bodyanskiy, Iryna Pliss, Olena Vynokurova

Control Systems Research Laboratory, Kharkiv National University of Radio Electronics, UKRAINE, Kharkiv, 14 Nauky Ave.,  
email: yevgeniy.bodyanskiy@nure.ua, iryna.pliss@nure.ua, vynokurova@gmail.com

**Abstract:** In this paper the autoencoder based on the generalized neo-fuzzy neurons is proposed. Also its fast learning algorithm was proposed. Such system can be used as part of deep learning neural networks. The proposed neo-fuzzy autoencoder is characterized by high learning speed and less number of tuned parameters in comparison with well-known approaches. The efficiency of proposed approach has been examined based on different benchmarks and real data sets.

**Keywords:** neo-fuzzy autoencoder, deep learning network, neo-fuzzy neuron, fast learning algorithm, data compression.

## I. INTRODUCTION

The task of compressing information that should be further processed, is one of main problem, which is solved in Data Mining. For solving this problem, a lot of approaches [1-5] are proposed, at that more of these methods comprehend the information processing in batch mode, when the fixed-sized data set is processed many times.

It is very important in order to the process of compression, the loss of information is minimal. Nowadays the approaches based on Deep neural networks (DNNs) [6-9] are widely used for solving the many tasks, which is connected with analysis of Big Data. As it can be seen from many researches the DNNs provide significantly better results than the conventional shallow neural networks.

The inherent part of DNN is, so-called autoencoder, which implements the compression of the input data and forms the input layers of the neural network.

As such autoencoders the multilayer associative “bottle-neck” perceptrons or restricted Boltzmann machines, in which nodes are the elementary Rosenblatt perceptrons with the sigmoidal activation functions, are often used.

Unfortunately, the learning process of such autoencoders demands a large time spending and cannot be implemented in online mode.

In the connection with the intensive development of Data Mining, Data Stream Mining, Web Mining over recent years the development of high speed information compression systems is an important problem. Such systems have to process data in sequential mode (perhaps in online mode) as the real information processing systems need.

## II. THE ARCHITECTURE OF NEO-FUZZY AUTOENCODER

Fig. 1 shows the architecture of the proposed autoencoder, which is autoassociative “bottle-neck” modification of

Kolmogorov’s neuro-fuzzy network [10-14] that implements the multiresolution approach and is the universal approximator according to the theorem of Kolmogorov-Arnold and Yam-Guen-Kreinovich.

It should be notice in [15, 16] the architecture, which nodes are the neo-fuzzy neurons (NFNs) [17], is considered. In spite of the simplicity of learning algorithm for synaptic weights, such system is abundant in the sense of the membership functions number.

Using the generalized neo-fuzzy neurons [18] instead of conventional NFN allows significantly to reduce the membership functions number and to introduce stacked NN [19]. Such stacked NN allows to simplify the architecture of autoencoder and in this way to speed up the learning process.

Therefore, autoencoder consists of two sequentially connected layers, which are implemented with the generalized neo-fuzzy neurons GNFN<sup>[1]</sup> and GNFN<sup>[2]</sup>.

The sequence of input signals, which have to compress  $x(k) = (x_1(k), x_2(k), \dots, x_n(k))^T \in R^n$ , ( $k = 1, 2, \dots$  is a number of the observation or a current instant of time), is fed to GNFN<sup>[1]</sup>.

GNFN<sup>[1]</sup> consists of the  $n$  multidimensional nonlinear synapses  $MNS_i^{[1]}$ ,  $i = 1, 2, \dots, n$ , where each of them has a one input,  $m$  outputs,  $h$  membership functions  $\mu_{ii}^{[1]}(x_i(k))$ ,  $i = 1, 2, \dots, h$  and  $mh$  tuned synaptic weights  $w_{ij}^{[1]}$ ,  $j = 1, 2, \dots, m$ .

The output of GNFN<sup>[1]</sup> is the compressed vector of the signals  $y(k) = (y_1(k), \dots, y_j(k), \dots, y_m(k))^T \in R^m$ ,  $m < n$ , which simultaneously is output of the autoencoder. The signal  $y(k)$  is fed to the inputs of GNFN<sup>[2]</sup>, which contains  $m$  inputs,  $m$  multidimensional nonlinear synapses  $MNS_j^{[2]}$ , where each of them has one input,  $n$  outputs,  $h$  membership functions  $\mu_{ij}^{[2]}(y_j(k))$ ,  $i = 1, 2, \dots, h$  and  $nh$  synaptic weights  $w_{ij}^{[2]}$ .

Thus, considered autoencoder contains  $2nmh$  tuned synaptic weights and  $(n+m)h$  membership functions that is significantly fewer than in the architecture in [20].

In the outputs of GNFN<sup>[2]</sup> the recovered signal  $\hat{x}(k) = (\hat{x}_1(k), \dots, \hat{x}_i(k), \dots, \hat{x}_n(k))^T$  is formed. In such manner the autoencoder is the autoassociative hybrid neo-fuzzy system of computational intelligence.

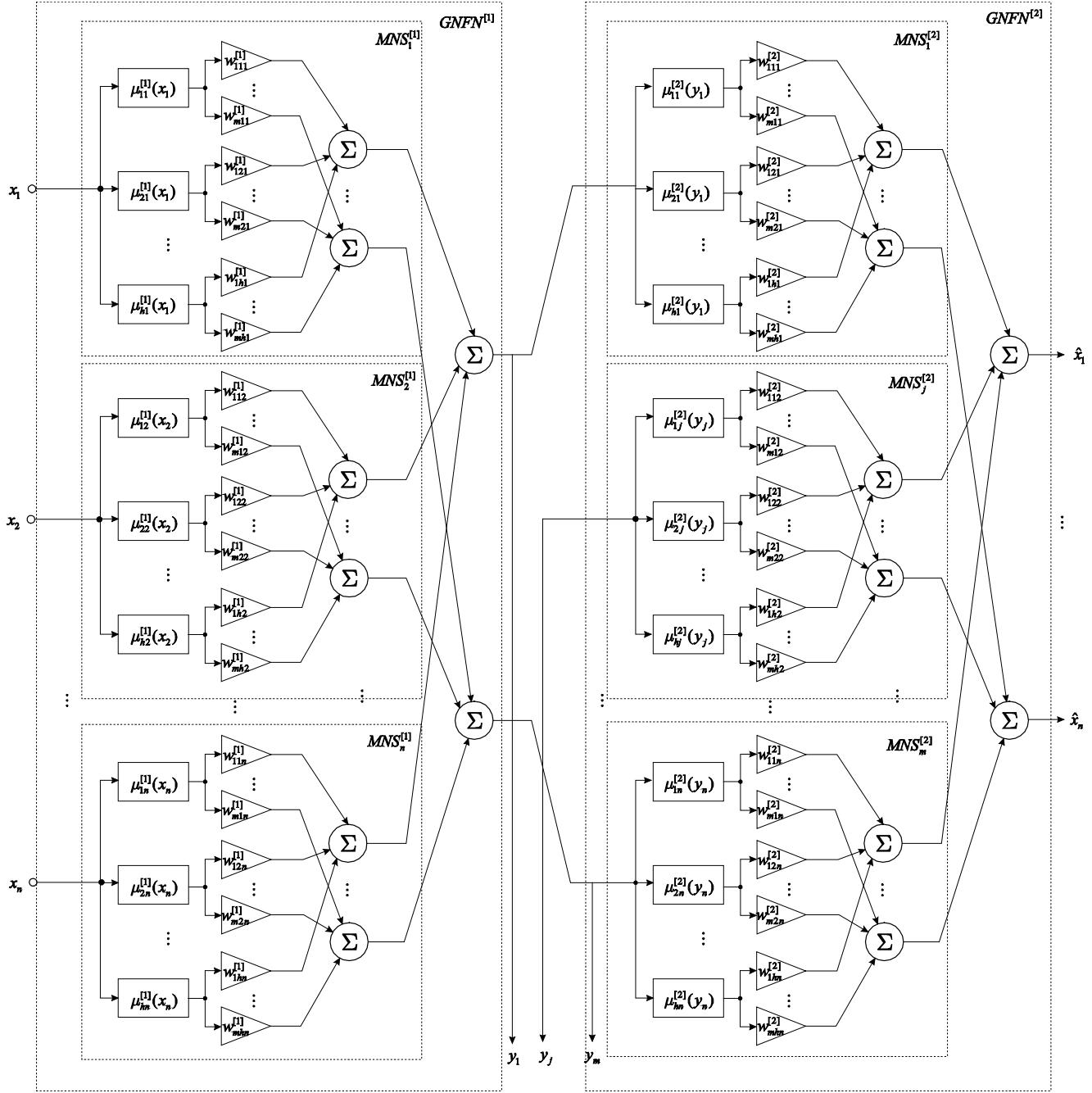


Fig 1. The architecture of the proposed neo-fuzzy autoencoder.

The proposed system is implemented a nonlinear mapping in the form

$$\hat{x}_i(k) = \sum_{j=1}^m \sum_{l=1}^h w_{ij}^{[2]} \mu_{lj}^{[2]} \left( \sum_{i=1}^n \sum_{l=1}^h w_{jli}^{[1]} \mu_{li}^{[1]}(x_i(k)) \right),$$

or in the matrix form

$$\hat{x}(k) = W^{[2]} \mu^{[2]}(W^{[1]} \mu^{[1]}(x(k))),$$

where

$$W^{[1]} = \begin{pmatrix} w_{111}^{[1]} & w_{121}^{[1]} & \dots & w_{1hn}^{[1]} \\ w_{211}^{[1]} & w_{221}^{[1]} & \dots & w_{2hn}^{[1]} \\ \vdots & \vdots & \ddots & \vdots \\ w_{m11}^{[1]} & w_{m21}^{[1]} & \dots & w_{mh1}^{[1]} \end{pmatrix},$$

$$W^{[2]} = \begin{pmatrix} w_{111}^{[2]} & w_{121}^{[2]} & \dots & w_{1hm}^{[2]} \\ w_{211}^{[2]} & w_{221}^{[2]} & \dots & w_{2hm}^{[2]} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n11}^{[2]} & w_{n21}^{[2]} & \dots & w_{nhm}^{[2]} \end{pmatrix},$$

$$\begin{aligned}\mu_{11}^{[1]}(x_1(k)) &= \left( \mu_{11}^{[1]}(x_1(k)), \mu_{21}^{[1]}(x_1(k)), \dots, \mu_{h1}^{[1]}(x_1(k)), \right. \\ \mu_{12}^{[1]}(x_2(k), \dots, \mu_{li}^{[1]}(x_i(k)), \dots, \mu_{hn}^{[1]}(x_n(k)) \left. \right)^T, \\ \mu_{11}^{[2]}(y_1(k)) &= \left( \mu_{11}^{[2]}(y_1(k)), \mu_{21}^{[2]}(y_1(k)), \dots, \mu_{h1}^{[2]}(y_1(k)), \right. \\ \mu_{12}^{[2]}(y_2(k), \dots, \mu_{lj}^{[2]}(y_j(k)), \dots, \mu_{hm}^{[2]}(y_m(k)) \left. \right)^T.\end{aligned}$$

### III. THE LEARNING ALGORITHM FOR SYNAPTIC WEIGHTS OF NEO-FUZZY AUTOENCODER

For the tuning the synaptic weights of GNFN<sup>[2]</sup> we can use the gradient procedure of minimizing the quadratic criterion in the form

$$\begin{aligned}w_{ij}^{[2]}(k) &= w_{ij}^{[2]}(k-1) - \eta^{[2]}(k) \frac{\partial e_i^2(k)}{\partial w_{ij}^{[2]}} = \\ &= w_{ij}^{[2]}(k-1) + \eta^{[2]}(k) e_i(k) \mu_{lj}^{[2]}(y_j(k))\end{aligned}$$

where  $\eta^{[2]}(k)$  is learning rate parameter of the output layer, which is chosen accordingly to the condition in [20, 21]

$$\eta^{[2]}(k) = (r^{[2]}(k))^{-1}; r^{[2]}(k) = \alpha r^{[2]}(k-1) + \|\mu^{[2]}(y(k))\|^2$$

where  $0 \leq \alpha \leq 1$  is forgetting factor.

For tuning the synaptic weights GNFN<sup>[1]</sup> the optimized backpropagation error procedure, which for uniformly distributed in the line of X-axis the triangular membership functions with centers  $\bar{x}_{ii}^{[1]} \bar{y}_{ij}^{[2]}$  can be write in the form

$$w_{ji}^{[1]}(k) = w_{ji}^{[1]}(k-1) + \eta^{[1]}(k) e_i(k) \mu_{ii}^{[1]}(x_i(k)) \tilde{w}_{ij}^{[1]}(k)$$

where

$$\begin{aligned}\eta^{[1]}(k) &= (r^{[1]}(k))^{-1}; r^{[1]}(k) = \alpha r^{[1]}(k-1) + \|\mu^{[1]}(x(k))\|^2, \\ \tilde{w}_{ij}^{[1]}(k) &= \sum_{l=1}^h w_{ij}^{[1]}(k) \begin{cases} (\bar{y}_{l,j}^{[2]} - \bar{y}_{l-1,j}^{[2]})^{-1}, & \text{if } y_j(k) \in [\bar{y}_{l-1,j}^{[2]}, \bar{y}_{l,j}^{[2]}], \\ (\bar{y}_{l,j}^{[2]} - \bar{y}_{l+1,j}^{[2]})^{-1}, & \text{if } y_j(k) \in [\bar{y}_{l,j}^{[2]}, \bar{y}_{l+1,j}^{[2]}], \\ 0 \text{ otherwise} \end{cases}.\end{aligned}$$

The proposed learning algorithm for synaptic weights of autoencoder is characterized by high speed and adjugate following and filtering properties.

### IV. EXPERIMENTS

For effectiveness verification of the proposed neo-fuzzy autoencoder, the data sets were taken from UCI Repository [22]: Iris, Wine, Hayes-roth. Data set "Iris" contains 150 observations (Number of Attributes: 4) of 3 classes, Data set "Wine" contains 178 observations (Number of Attributes: 13) of 3 classes, data set "Hayes-roth" contains 160 observations (Number of Attributes: 5) of 3 classes.

It is seen from Fig.2 data, which are compressed using neo-fuzzy autoencoder, are more compact clusters than data, which

are compressed based on the autoassociative multilayer neural network "Bottle Neck".

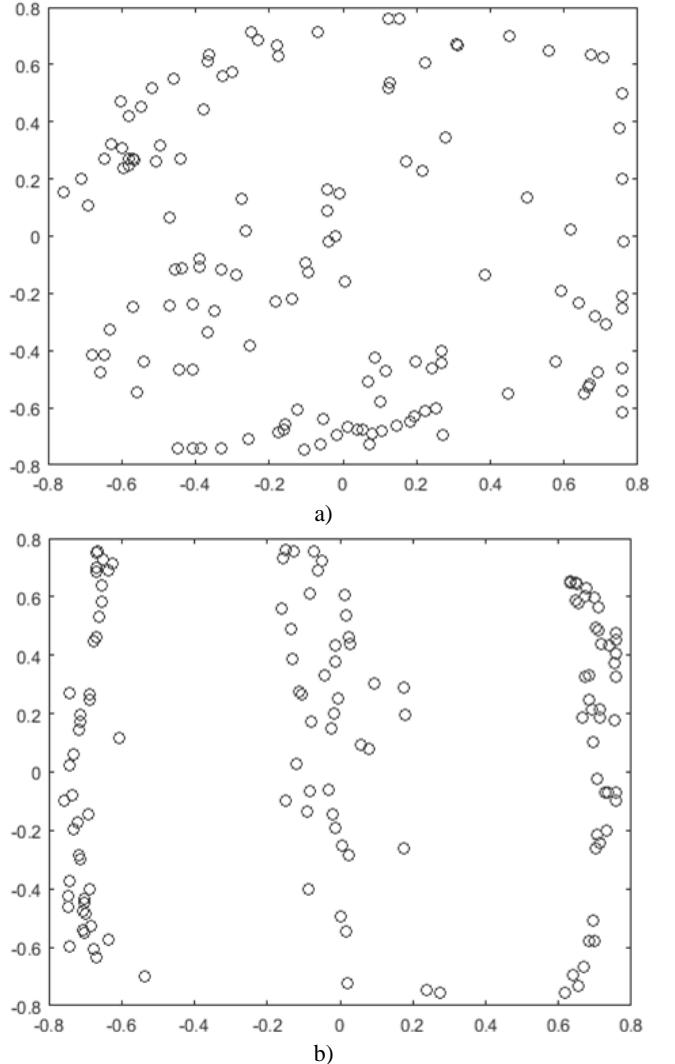


Fig. 2. Data set Hayes-roth after compression based on the autoassociative multilayer neural network "Bottle Neck" (a) and neo-fuzzy autoencoder (b)

The results, which were obtained using the proposed neo-fuzzy autoencoder, were compared with the results of autoassociative multilayer neural network "Bottle Neck" (Table I). The dimension of compression data was 2 components. The simulation was performed 20 times with different initial condition and the results were averaged.

TABLE I. RESULTS OF SIMULATION

| AUTOENCODERS   | DATA SETS  | MSE   |
|--|------------|-------|
| Neo-fuzzy autoencoder                                    | Iris       | 0.199 |
|  | Wine       | 0.499 |
|  | Hayes-roth | 0.312 |
| Autoassociative three layer neural network "Bottle Neck" | Iris       | 0.486 |
|  | Wine       | 0.903 |
|  | Hayes-roth | 0.593 |

#### IV. CONCLUSIONS

The architecture of «bottle-neck» two-layer autoencoder and its learning algorithm are proposed. Such system is based on generalized neo-fuzzy neurons and is autoassociative “bottle-neck” modification of Kolmogorov’s neuro-fuzzy network. The proposed hybrid neo-fuzzy system of computational intelligence provides high quality of information compression, which are fed sequentially for processing. It is characterized by computational simplicity and high speed of the learning process.

#### REFERENCES

- [1] J. Han and M. Kamber, “Data Mining: Concepts and Techniques”. Amsterdam: Morgan Kaufman Publ., 2006.
- [2] C.C. Aggarwal, “Data Mining”, N.Y.: Springer, 2015.
- [3] A. Bifet, R. Gavaldà, G. Holmes, and B. Pfahringer, Machine Learning for Data Streams with Practical Examples in MOA. The MIT Press, 2018.
- [4] A. Bifet, Adaptive Stream Mining: Pattern Learning and Mining from Evolving Data Streams. Amsterdam: IOS Press, 2010.
- [5] A. Menshawy, Deep Learning By Example: A hands-on guide to implementing advanced machine learning algorithms and neural networks. Packt Publishing Limited, 2018.
- [6] M. Fullan, J. Quinn, and J. McEachen, Deep Learning: Engage the World Change the World. Corwin, 2017.
- [7] A. L. Caterini and D. E. Chang, Deep Neural Networks in a Mathematical Framework. Springer, 2018.
- [8] Y. LeCun, Y. Bengio, and G.E. Hinton, “Deep Learning”. *Nature*, 2015, v. 521, pp. 436-444.
- [9] D. Graupe, “Deep Learning Neural Networks: Design and Case Studies”. World Scientific Publishing Company, 2016.
- [10] V. Kolodyazhniy and Ye. Bodyanskiy, “Fuzzy Kolmogorov’s Network,” in *Lecture Notes in Computer Science*, vol. 3214, M.G. Negoita et al., Eds., Springer-Verlag, 2004, pp.764-771.
- [11] Ye. Bodyanskiy, V. Kolodyazhniy and P. Otto, “Neuro-fuzzy Kolmogorov’s network for time-series prediction and pattern classification,” in *Lecture Notes in Artificial Intelligence*, vol. 3698, U. Furbach, Ed., Heidelberg: Springer –Verlag, 2005, pp. 191-202.
- [12] V. Kolodyazhniy, Ye. Bodyanskiy and P. Otto, “Universal approximator employing neo-fuzzy neurons,” in *Computational Intelligence Theory and Applications*, Ed. B. Reusch, Ed., Berlin-Heidelberg: Springer, 2005, pp. 631-640.
- [13] V. Kolodyazhniy, Ye. Bodyanskiy, V. Poyedyntseva, and A. Stephan “Neuro-fuzzy Kolmogorov’s network with a modified perceptron learning rule for classification problems,” in *Advances in Soft Computing*, vol. 38, B. Reuch, Ed., Berlin-Heidelberg: Springer-Verlag, 2006, pp. 41-49.
- [14] Ye. Bodyanskiy, Ye. Gorshkov, V. Kolodyazhniy, and V. Poyedyntseva “Neuro-fuzzy Kolmogorov’s network,” in *Lecture Notes in Computer Science*, vol.3697, W. Duch, J. Kacprzyk, E. Oja, and S. Zadrożny, Eds., Berlin-Heidelberg: Springer-Verlag, 2005, pp.1-6.
- [15] V. Kolodyazhniy, F. Klawonn, and K. Tschumitschew, “A neuro-fuzzy model for dimensionality reduction and its application” *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* vol. 15, is. 05, October 2007, pp. 571-593.
- [16] Vynokurova O., Bodyanskiy Ye., Pliss I., Peleshko D., Rashkevych Yu. “Neo-fuzzy encoder and its adaptive learning for Big Data processing.” *Scientific Journal of RTU, Series “Computer Science”* Volume “Information Technology and Management Science” 2017, vol. 20, pp. 6–11.
- [17] T. Yamakawa, E. Uchino, T. Miki and H. Kusanagi, “A neo-fuzzy neuron and its applications to system identification and prediction of the system behavior,” in *Proc. of 2-nd Int. Conf. on Fuzzy Logic and Neural Networks “IIZUKA-92”*, Iizuka, Japan, pp. 477–483, 1992.
- [18] R.P.Landim, B. Rodrigues, S.R. Silva, and W.M. Caminhas, “A neo-fuzzy-neuron with real time training applied to flux observer for an induction motor”. In: *Proceedings of IEEE Vth Brazilian Symposium on Neural Networks, Belo Horizonte*, 9-11 Dec 1998, pp. 67-72.
- [19] J. Schmidhuber, “Deep learning in neural networks: An overview,” *Neural Networks*, vol. 61, pp. 85-117, Jan. 2015. (doi: 10.1016/j.neunet.2014.09.003)
- [20] Ye. Bodyanskiy, I. Kokshenev, V. Kolodyazhniy, “An adaptive learning algorithm for a neo fuzzy neuron,” in *Proc. 3rd Int. Conf. of European Union Society for Fuzzy Logic and Technology (EUSFLAT 2003)*, Zittau, 2003, pp. 375-379.
- [21] P. Otto, Ye. Bodyanskiy, V. Kolodyazhniy, “A new learning algorithm for a forecasting neuro-fuzzy network,” *Integrated Computer-Aided Engineering*, vol. 10, pp. 399-409, Dec. 2003
- [22] UCI Repository of machine learning databases. CA: University of California, Department of Information and Computer Science. [Online]. Available: <http://www.ics.uci.edu/~mlearn/MLRepository.html>

# Software Engineering

# Intelligent System Analyzing Quality of Land Plots

Andrii Bayurskii<sup>1</sup>, Svitlana Krepuch<sup>2</sup>

1. Department of Computer Science, Ternopil National Economic University, UKRAINE, Ternopil, 8 Chekhova str., email: andriy.bayurkiy@gmail.com

2. Department of Computer Science, Ternopil National Economic University, UKRAINE, Ternopil, 8 Chekhova str., email: s.krepuch@tneu.edu.ua

**Abstract:** Products grown on the territory of Ukraine or goods obtained from the processing of the crops concerned are attractive in many countries of the world. It is advisable to focus on improving the cultivation of crops to increase the export of our goods abroad. The development of an appropriate software complex that would combine the interests of the state and farmers to increase the yield of crops and, accordingly, increase the amount of crop, is an integral part of improving the work of the agricultural sector of the state. Introduction a crop rotation is the main idea of developing a software system. This will lessen the drying of the soil and increase yield.

**Keywords:** mathematical model, forecasted yield, crop rotation, land plot, agrarian.

## I. INTRODUCTION

Each year the state must sell the tender for the planting of certain crops for a certain period of time. The most effective distribution of orders for planting between farmers is the main issue of study. There is a problem in absence of a database of agrarians and information's about tenders at the first stage. In the work, we will use a simulated bidding scheme based on input data that is as close as possible to the actual data.

Only a few hundred species of plants are used in agriculture. Most of the agricultural products are provided at best twenty species of plants. For thousands of years, people have been trying to improve their yields, using the remains of fauna and flora. However, it was a natural way to improve yields. Today we see a completely different picture. On the one hand it increases yields but on the other hand the quality of crops and also the ecological condition of soils and groundwater deteriorate [1].

Considering all the above problems, the creation of an intelligent system analyzing quality of land plots will be a major breakthrough in the field of agronomy. It will allow maximize profits for both the state and agrarians by efficient cultivation of different crop.

## II. MODELING THE PROCESS OF YIELD OF CULTURE

The main idea of developing a software system for planting land is to improve soil fertility through such a crop rotation of crops. This will allow to reduce drying soil, increase the yield of the crop and will require less use of various types of herbicides, pesticides, supplements, growth

stimulants, etc. At the end, it will still have a positive impact on the environment, because less chemicals will fall into the soil and, accordingly, to the reservoirs.

Another positive aspect is that such a system does not exist in Ukraine today. The main emphasis of the state's provision of export goods in the field of agrarian industry is focused on agrarians-monopolists. They conclude with state the main tendering agreements. Within these agreements, the main focus is on obtaining a stable yield, resulting in a soil deterioration. It leads to extensive use of a large amount of inorganic fertilizers in order to improve soils quality which in turn adversely affects the environmental state or condition.

The main crops grown in the Ternopil region are sunflowers, corn, rape, sugar beet and winter crops - wheat and barley. These crops are an export commodity within our state, and some of them, including sunflowers and wheat, are part of the exported goods abroad.

The average yield index of the described crops for Ternopil region in 2016 was:

- 3.8 t from 1 ha for wheat;
- 2.4 t from 1 ha for sunflowers;
- 30 t from 1 ha for sugar beet;
- 4.7 t from 1 ha for corn;
- 3.6 t from 1 ha for barley;
- 1.8 t from 1 ha for rape.

Let's take these indicators as etalon for standard crop yields to determine which crop on which area is better to plant to obtain the best possible yield without excess fertilizer. The planting of a culture will then influence what culture was planted in this field last year and how many years the proposed culture was not grown in this area.

The most depleting for soil, from the crops grown in the Ternopil region are sunflowers and corn. The land should reset for seven years to recapture the previous fertilizing power and inspiration after the sunflowers, after the corn only three years. This term agricultural producers do not always stand, or do not stand at all. This is especially true for sunflower. So in the system is proposed to consider planting a plot sunflower at least after 4 years.

Table 1 summarizes the planting rates of crops. It is including the last year of landing and crop rotation, as well as the obtained average yield of a crop from 1 ha of land. Let's enter some variables:

- $y$  – the maximum number of years the culture should not be grown on the site;
- $v_{pr_i}$  - average yield of the crop that was last planted on the plot within the proposed area for the last landing;
- $v_i$  - the average yield of a crop that is being prepared for planting on a plot within the proposed area for the last landing;

-  $v_{res_i}$  - average yield of cultivated crop.

TABLE 1. INDICATORS OF AVERAGE CROP YIELDS ACCORDING TO THE PREVIOUS CROP ROTATION AND LAST YEAR OF CULTIVATION

| Culture type | $y$ | $v_{pri}$ | $v_i$ | $v_{res_i}$ |
|--------------|-----|-----------|-------|-------------|
| Sunflower    | 3   | 4.7       | 2.4   | 2.2         |
|              | 3   | 3.6       | 2.4   | 2.4         |
|              | 3   | 1.8       | 2.4   | 2.5         |
|              | 3   | 3.8       | 2.4   | 2.3         |
|              | 3   | 30        | 2.4   | 2.6         |
| Corn         | 2   | 3.4       | 4.7   | 4.4         |
|              | 2   | 3.6       | 4.7   | 4.7         |
|              | 2   | 1.8       | 4.7   | 4.8         |
|              | 2   | 3.8       | 4.7   | 4.7         |
|              | 2   | 30        | 4.7   | 4.8         |
| Barley       | 1   | 4.7       | 3.6   | 3.5         |
|              | 1   | 2.4       | 3.6   | 3.4         |
|              | 1   | 1.8       | 3.6   | 3.7         |
|              | 1   | 3.8       | 3.6   | 3.6         |
|              | 1   | 30        | 3.6   | 3.7         |
| Rape         | 1   | 4.7       | 1.8   | 1.7         |
|              | 1   | 2.4       | 1.8   | 1.6         |
|              | 1   | 3.6       | 1.8   | 1.8         |
|              | 1   | 3.8       | 1.8   | 1.8         |
|              | 1   | 30        | 1.8   | 1.9         |
| Wheat        | 1   | 4.7       | 3.8   | 3.7         |
|              | 1   | 2.4       | 3.8   | 3.6         |
|              | 1   | 3.6       | 3.8   | 3.8         |
|              | 1   | 1.8       | 3.8   | 3.9         |
|              | 1   | 30        | 3.8   | 4           |
| Sugar beet   | 1   | 4.7       | 30    | 28          |
|              | 1   | 2.4       | 30    | 27          |
|              | 1   | 3.6       | 30    | 30          |
|              | 1   | 1.8       | 30    | 32          |
|              | 1   | 3.8       | 30    | 31          |

It is advisable to construct a model of the dependence of the yield of the crop from its crop rotation within a specific area. At the initial stage, we will choose the linear structure of the model of the species:

$$v_{res_i} = k_1 \cdot y_i + k_2 \cdot v_{pri} + k_3 \cdot v_i + k_4 \quad (1)$$

where  $k_i$ ,  $i = \overline{1..4}$  - unknown coefficients, the values of which need to be evaluated on the basis of the analysis of the data presented in Table 1.

According to tabular data we will make a system of linear algebraic equations in this form:

$$\begin{cases} k_1 * y_i + k_2 * v_{pri} + k_3 * v_i + k_4 = v_{resi} \\ \dots \\ k_1 * y_i + k_2 * v_{pri} + k_3 * v_i + k_4 = v_{resi} \\ \dots \\ k_1 * y_{30} + k_2 * v_{pri30} + k_3 * v_{i30} + k_4 = v_{res30} \end{cases} \quad (2)$$

The solution of the system (2) is the region of coefficients of the model [2,3]. Using the least squares method to find the estimates of the coefficients of the model from SLR (2), we obtain the following model:

$$v_{resi} = 0,008 * y_i + 0,006 * v_{pri} + 0,99 * v_i - 0,04 \quad (3)$$

Obtained value from the formula (3) is reflects the average yield of the proposed landing in the specified section of the crop with regard to the history of the planting. However, for a more adequate construction of the planting strategy, it is necessary to take into account the risks that can reduce this indicator.

Let's establish a 5% deviation from the predicted yield index for the possibility of taking into account the various risks of deteriorating crop yields. As a result, the forecast yield of the crop will be in the range [3]:

$$[v_{resi}^-; v_{resi}^+] = [v_{resi} - 0,05 \cdot v_{resi}; v_{resi}] \quad (4)$$

The order of the state to a certain culture appears in the interval:

$$[V_{con_i}^-; V_{con_i}^+] \quad (5)$$

where  $V_{con_i}^-$  - the minimum amount of crop yield required by the state to cover the domestic needs and export needs of the overseas;  $V_{con_i}^+$  - the maximum amount of crop yield that the state may buy overtime from agrarians.

The function of the distribution of planting the culture between the plots within a single order (2.5) is represented by the expression:

$$[f(V_i)] = \sum_{i=1}^{S_i} (S_i \cdot [v_{resi}^-; v_{resi}^+]), \quad (6)$$

where  $S_i$  - the area of the field selected for planting the selected culture on the  $i$ -th iteration.

The number of iterations is determined in accordance with the improvement of the goal function  $F_i$ , the value of which is determined by the formula:

$$F_i = \min_{i=1, \dots, N} \{mid([V_{con_i}^-; V_{con_i}^+]) - mid([f(V_i)])\} \quad (7)$$

where  $F_i \geq 0$ .

That is, the choice of the optimal field for planting will be selected as long as the function  $F_i$  will accept positive values from a set of possible values [4].

### III. SOFTWARE REALIZATION

Figure 3.1 depicts a general block diagram of the system [5,6].

Basically you can highlight three main stages of the program system:

1) Formation of an order by the state. This process includes selecting the necessary crops, sorting them by priority and indicating the interval of the yield. Flow chart of this process depicted in Fig 3.

2) Planting cost estimation. The software system analyzes the history of the planting of all available fields of the agrarians on which it is planned to plant the corresponding crop and publish a list of fields with the most effective way of planting, that is, those fields that will result in the best crop yields taking into account crop rotation. Flow chart of this process depicted in Fig 2.

3) Giving of a result. The last step will be the window with the formed fields for planting each crop. At this stage, you can review the forecast average yield, previous yields in this field and other indicators. Flow chart [7] of this process depicted in Fig 4.

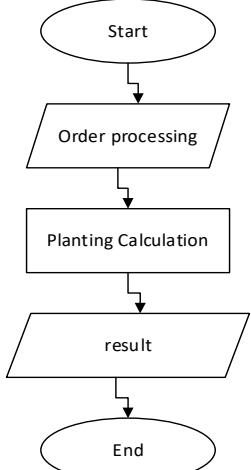


Fig. 1. Flow chart diagram

A detailed flowchart of the "Planting Calculation" process is shown in Fig. 2. Consider the step-by-step description of the process:

1. System choose the culture from order. The culture was sorted by priority.
2. Retrieve information about all available fields on which can be planted the culture, selected in step 1.
3. Calculating for each field the estimated yield of the selected crop depending on the history of planting and the average yield in this field by the formula 4.
4. Select the field with the largest interval of the forecast yield according to the formula 6.
5. Connecting the field to the chosen culture.
6. This field will be marked as selected and will not appear in field list by the end of calculation.
7. Check whether there any culture that does not connected to any of fields;
- 7.1. If fields are selected for all cultures, then proceed to the next stage;

- 7.2. If there are cultures for which the fields are not yet selected, then choose the culture that follows the priority and proceed to step 2.

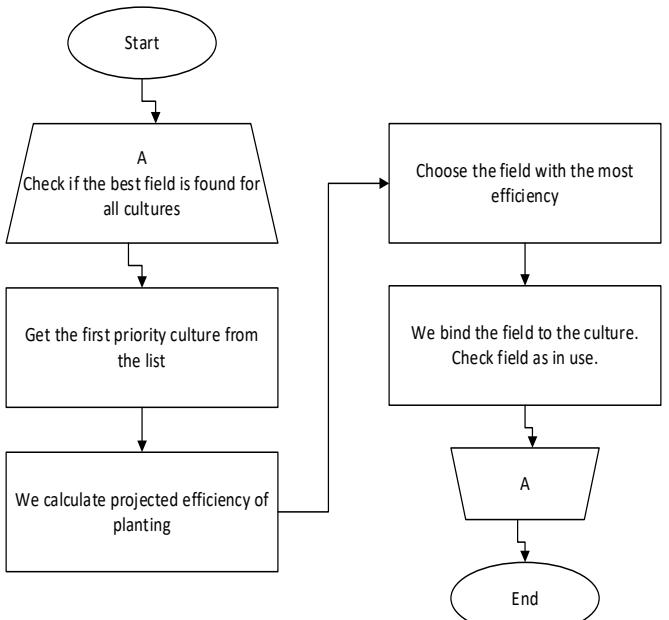


Fig. 2. Flow chart of the «Planting Calculation» process

In order to get started with the system, it is necessary to form orders. We should choose the culture, their priority and indicate the minimum and maximum crop yields for the state. The form of this process is depicted in Fig. 3. [7]

| OrderForm   |           |           |         |         |           |           |       |     |     |           |     |     |            |     |     |      |     |     |        |     |     |      |     |     |
|---|-----------|-----------|---------|---------|-----------|-----------|-------|-----|-----|-----------|-----|-----|------------|-----|-----|------|-----|-----|--------|-----|-----|------|-----|-----|
| Culture   |           | Add       | Perform |         |           |           |       |     |     |           |     |     |            |     |     |      |     |     |        |     |     |      |     |     |
| Rape  |           | Add       | Perform |         |           |           |       |     |     |           |     |     |            |     |     |      |     |     |        |     |     |      |     |     |
| Open results  |           |           |         |         |           |           |       |     |     |           |     |     |            |     |     |      |     |     |        |     |     |      |     |     |
| <table border="1"> <thead> <tr> <th>Culture</th> <th>min yield</th> <th>max yield</th> </tr> </thead> <tbody> <tr><td>Wheat</td><td>100</td><td>120</td></tr> <tr><td>Sunflower</td><td>120</td><td>140</td></tr> <tr><td>Sugar beet</td><td>140</td><td>160</td></tr> <tr><td>Corn</td><td>160</td><td>180</td></tr> <tr><td>Barley</td><td>180</td><td>200</td></tr> <tr><td>Rape</td><td>200</td><td>220</td></tr> </tbody> </table> |           |           |         | Culture | min yield | max yield | Wheat | 100 | 120 | Sunflower | 120 | 140 | Sugar beet | 140 | 160 | Corn | 160 | 180 | Barley | 180 | 200 | Rape | 200 | 220 |
| Culture   | min yield | max yield |         |         |           |           |       |     |     |           |     |     |            |     |     |      |     |     |        |     |     |      |     |     |
| Wheat   | 100       | 120       |         |         |           |           |       |     |     |           |     |     |            |     |     |      |     |     |        |     |     |      |     |     |
| Sunflower   | 120       | 140       |         |         |           |           |       |     |     |           |     |     |            |     |     |      |     |     |        |     |     |      |     |     |
| Sugar beet  | 140       | 160       |         |         |           |           |       |     |     |           |     |     |            |     |     |      |     |     |        |     |     |      |     |     |
| Corn  | 160       | 180       |         |         |           |           |       |     |     |           |     |     |            |     |     |      |     |     |        |     |     |      |     |     |
| Barley  | 180       | 200       |         |         |           |           |       |     |     |           |     |     |            |     |     |      |     |     |        |     |     |      |     |     |
| Rape  | 200       | 220       |         |         |           |           |       |     |     |           |     |     |            |     |     |      |     |     |        |     |     |      |     |     |

Fig. 3. Purchase Order Form

Once the system has selected the most effective options for planting fields, you can open the resulting form for viewing detailed information on the results of calculations, namely:

- description of the data on the agrarian who owns the field with the planted crop;
- field area;
- a culture that will be planted in the chosen field;
- current year of planting.

The prototype of the shape is depicted in Fig. 4.

| ResultViewForm     |       |            |      |                                     |
|--------------------|-------|------------|------|-------------------------------------|
| Agrarian           | Field | Culture    | Year | Done                                |
| Maria              | 20 ha | Wheat      | 2017 | <input checked="" type="checkbox"/> |
| Maria              | 30 ha | Sunflower  | 2017 | <input checked="" type="checkbox"/> |
| Good person        | 25 ha | Sugar beet | 2017 | <input checked="" type="checkbox"/> |
| Private agrarian 1 | 25 ha | Corn       | 2017 | <input checked="" type="checkbox"/> |
| Maria              | 35 ha | Barley     | 2017 | <input checked="" type="checkbox"/> |
| Good person        | 40 ha | Rape       | 2017 | <input checked="" type="checkbox"/> |

Fig. 4. Result form

In general, this form depicts a compiled information about agrarian and field, which will effectively be used when planting a certain culture. Double-clicking on a culture will open a dialog box that displays detailed information about the field and the owner. The prototype of the shape is depicted in Fig. 5.

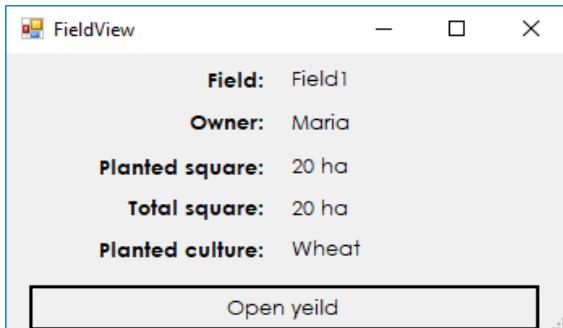


Fig. 5. View field information

Clicking on the "View yield" button will open a window that displays information about how much culture of the whole field was collected, how much was forecasted for collection, the average yield of actual and average yields is projected in the range from the minimum to the maximum value. The prototype of the shape is depicted in Fig. 6.

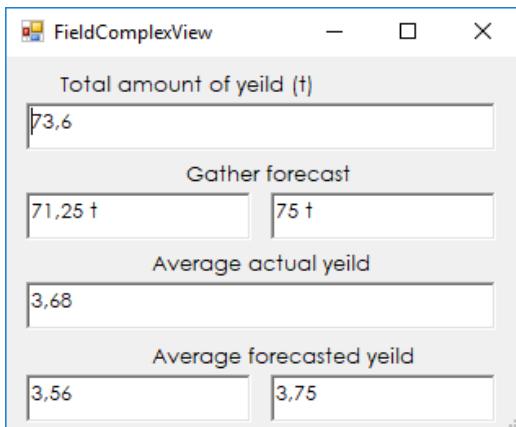


Fig. 6. Display detailed information on crop yields at a particular site

#### IV. CONCLUSION

In the scientific work the actual scientific research task of analyzing the quality of land plot planting has been solved. The following scientific and practical results were obtained.

During the analysis of the subject area it was discovered that today the main emphasis of the state's provision of export goods in the field of agrarian industry is focused on agrarians of the monopolists. Between them and the state, the main

tendering agreements are concluded. Within these agreements, the main focus is on obtaining a maximum yield, which leads to a deterioration of the soil, which is trying to improve with the use of a large amount of fertilizers of inorganic origin, which in turn adversely affects the environmental state of the environment.

In order to improve crop yields and reduce the amount of inorganic fertilizers, it is proposed to take into account the previous crop rotation when planting the crop, which can significantly improve the yields of the proposed crop.

In the framework of the proposed approach, a mathematical model of the dependence of the crop yield from the previous crop rotation within a specific area is developed.

The method of planting of plots on the basis of the developed model of crop rotation is developed, which will allow covering the minimum amount of crop yield required for the state for crops and improve the quality of soil application and, as a result, average yields of crops.

The program complex of quality of planting of land plots has been developed, which will improve the connection between "state-agrarian state". Provide the opportunity for private entrepreneurs to enter a large market of sales and as a result, with the correct and lawful application of the software complex to improve the quality of land and the ecological state of the country.

#### REFERENCES

- [1] "Strategy for the development of the agrarian sector of the Ukrainian economy for the period up to 2020," 2013. [Online]. Available at: <http://minagro.gov.ua/node/7644>. [Application Date: 08.12.2017 November 2017].
- [2] S. Ya. Krepych "Modeling and providing of functional suitability of static systems by methods of analysis of interval data", *PhD thesis*, Lviv, 2016, 166 pp. (in Ukrainian)
- [3] S. Ya. Krepych "Comparative analysis of the time complexity of random searches in the synthesis problem with given allowable values of output characteristics and tolerances on the parameters of its elements" *TNTU Bulletin, Scientific Journal*, 2015, No. 1, pp. 204-219. (in Ukrainian)
- [4] G. Alefeld, Yu Herzberger: "Introduction to interval calculations" Moscow: World, 1987, 360 p. (in Russian)
- [5] Chonoles M. J., Schardt J.A. UML 2 for Dummies. – Hungry Minds, 2003, 412 pp. (in Russian)
- [6] Kendal S. Fast Track UML 2.0. Apress, 2004, 416 pp.
- [7] Skiena, S., 2013. The Algorithm Design Manual. 2nd ed. Springer Publishing Company: Springer.

# Software Based on Blockchain Technology for Consolidation the Medical Data about the Patients Examination

Andriy Pukas, Vitalii Smal, Vadym Zabchuk

Department of Computer Science, Ternopil National Economic University, UKRAINE, Ternopil, 8 Chekhova str.,  
email: [apu@tneu.edu.ua](mailto:apu@tneu.edu.ua), [vitalii.smal@gmail.com](mailto:vitalii.smal@gmail.com), [vadzab5@gmail.com](mailto:vadzab5@gmail.com)

**Abstract:** Software architecture based on blockchain technology for efficient medical records exchanging in safe mode, data confidentiality protecting, and giving to patients more control over their medical information is described in this paper. New software system architecture was proposed to consolidate medical data on patient screening based on blockchain technology and decentralization principles, which increased the level of data access security in heterogeneous medical systems.

**Keywords:** software, data consolidation, blockchain, patient examination.

## I. INTRODUCTION

Hospital-based medical information systems are becoming increasingly popular worldwide. Paper medical cards are gradually losing their relevance, being replaced by hospital-based medical information systems. Generally doctors are used them and already know how to use them properly. Such systems have many advantages. First, the electronic card will never get lost, and the patient will not be able to take it home. Thus, the information is always kept in medical facility. The second advantage of an e-card is that there is no need to look for it to be later handed over to an appropriate specialist by the receptionist. All data is always available on a doctor's computer. Another advantage is that it eliminates the need for a permanent patching of additional sheets, advisory opinions, analyses and research results. All such information is recorded into the specified fields of the program, which gives the necessary information at doctor's first request. It also enables several specialists of the clinic to simultaneously get access to patient's electronic card contents. It makes possible for doctors not only to simultaneously read the patient's card, but also to fill it. This feature significantly optimizes the medical staff activities.

Despite the advantages, they also have drawbacks. They are very fragmented and decentralized. Another problem is that such systems are not unified and different medical institutions use different means of electronic data storage. This significantly complicates the transfer of such data between healthcare facilities. Consider a case when a patient needs to show their disease history to a doctor of another medical establishment. To do this, the patient has to make a disease history request and wait for it to be processed. It can take a long time, and in some situations, even a few hours can significantly affect the person's health or even life.

Production engineers and healthcare professionals [1, 2] worldwide see blockchain technologies [3] as a way to efficiently exchange medical records in safe mode, protect the data privacy from hackers, and give patients more control over their information.

The development of a blockchain-based system that would consolidate patient medical data from different providers should address such issues as fragmentary and slow access to patient medical data, incompatibility of medical information systems, and improve the quality and quantity of data for medical research.

## II. TASK STATEMENT

The aim of research is to consolidate medical data on patient surveys from various medical institutions and increase the level of access security to such data, based on creation a prototype of software system built on principles of decentralization and blockchain architecture.

Research objectives:

- analysis of modern electronic medical systems;
- design a secure and compatible software system for the medical data consolidation;
- implementation a prototype of software system based on blockchain technology.

It is easy to assume that when a doctor examines a patient or gives them a new prescription, the patient will agree to add a reference or "index" to blockchain - a decentralized electronic system similar to the bitcoin [4]. But instead of making payments, this block chain will write medical information in a cryptographic database supported by a network of computers that is accessible to anyone who works with it. Each index that the physician will add to the journal will become part of the patient's registry, regardless of which electronic system the doctor used. Therefore, any other doctor will be able to use it without worrying about the issues of incompatibility.

## III. ANALYSIS OF MEDICAL INFORMATION SYSTEMS IN UKRAINE

Electronic medicine is a development synthesis of medical and information technologies. This trend consists of many areas: from the creation of open digital registries of patients to their remote treatment.

The following are major factors in the field of electronic medicine development:

- introduction of automated informational sectoral systems, which, in particular, will enable the transition to electronic processing of medical documentation;
- development of telemedicine;
- improvement of the national health monitoring system;
- creation and implementation of new computer technologies of disease prevention, diagnostics, medical process support;
- creation of publicly available electronic medical resources;
- development of self-diagnostic methods and the construction of a personal health paradigm by e-medicine methods.

Information technology today can make medicine more affordable and effective. Many electronic systems exist on the Ukrainian market: Helsi, MIS EMSIMED, Doctor Eleks, MEDSTAR, MEDICS, Queue-Free Clinic etc [5].

All of the above systems are good solutions that are already implemented and work successfully in health care facilities. However, lack of properly established connection between them appears to be their main disadvantage. Each health care facility has its own system, so when a patient is treated in two or more different medical facilities, that person's analyses, examinations, and histories of illness are scattered across different medical systems. The patient does not have a single safe place, protected from hackers, to store their medical data, consolidated from different sources.

Therefore, having analyzed the Ukrainian medical systems market, we found the need for a software system that could consolidate medical data from various electronic medical systems in Ukraine and meet requirements such as reliability, security, and intuitive interface.

#### IV. ARCHITECTURE OF SOFTWARE SYSTEM

A software system for consolidating medical data is based on Ethereum [6], a platform for creating decentralized online blockchain services based on intelligent contracts [7]. Intelligent contracts are scripts that simplify, verify, ensure negotiations, execution of a contract, or check unwanted clauses of the agreement. Intelligent contracts, as a rule, also have a user interface and often follow the logic of contractual provisions. Thus, intelligent contracts allow for more complex blockchain transactions. Ethereum consists of a system of nodes (personal computers, clusters, virtual machines) in a decentralized network. Smart contracts are not a substitute for contracts in the traditional sense, but act as agreements on the implementation of certain actions or code. In this case, these contracts can be used to encode a set of indexes to medical data placements.

Lets consider the structure of smart contracts. The proposed system does not store electronic medical cards directly on Ethereum, but instead uses a relational set of smart contracts to encode indexes that can be used to locate and authenticate to the medical cards storage point. The system identifies three main types of contracts owned by patients, suppliers and other consumers. Namely:

- Registration contract;
- Patient-provider contract;
- Final contract.

The registration contract reflects the accordance of the participants' identifiers (patients, providers) with the Ethereum address (equivalent to the public key). The

regulation of new identities can be encoded in the contract, ensuring that only certified institutions can add new information to the blockchain. In turn, new information about the patient (for example, about new relationships) is added only with the approval of this patient. Each identification string is located at its blockchain address, where it is referred to by the final contract.

A contract on relationship between the patient and the provider is concluded between the two nodes of the system, where one node stores and manages medical records for another. Although we consider the case of a patient getting medical care at the healthcare institution, this concept applies to any pairwise interaction with data support. The patient-provider relationship defines the range of data indexes and associated access permissions that identify the records stored by the care provider. Each index consists of a query string, which, when executed in the supplier database, returns a subset of the patient data. The request string/line is embedded in the hash of this data subset to ensure that the data has not been changed in the source. Additional information shows where one can get access to the provider's database in the network, for example, the host name and port in the standard network topology. Data requests and related information are developed by the care provider and modified when new records are added. To allow patients to share records with other users, the dictionary implementation (hash table) displays the addresses of users of the list of additional request lines. Each line may indicate the portion of the patient's data to which the third-party user has access permission.

A prototype created demonstrates this design with SQL data queries. In a simple situation, the supplier refers to the patient's data by a simple SELECT request, based on the patient's address. Patients can use a tool that allows them to check the fields they want to share through the developed graphical interface. The system formulates the corresponding SQL queries and downloads them to the patient-provider relationship contract in a particular block. It's worth noting that, using common lines, the system can closely interact with any database implementation. Consequently, the prototype can be conveniently integrated with the existing infrastructure for data storage of the provider. At the same time, patients engage their micro level control of access to their medical records, thus choosing any part they want to share.

The final contract is comparable to the bread crumble trail, where each participant can find a summary of their relationship with any other participant. The final contract encodes the list of links to contracts on relations between patients and suppliers, providing for both current and previous interactions with other nodes of the system. Each relationship also stores the "status" variable indicating when the relationship was established and whether it was approved by the patient. Acceptance, rejection or removal of the relationship is controlled by the patient, giving full control over all the records in their history that they want to acknowledge. This function of the system is the key to satisfying its convenience criterion: an index to fragmented records is made in a single dedicated location.

It is shown in Fig. 1 the possible connections between different contracts and between customers and suppliers. It is

worth noting that the variable of the status of a particular contract may have different values depending on the permissions that it permits. Contracts are also used only for indexes: database requests that return records are processed off-line.

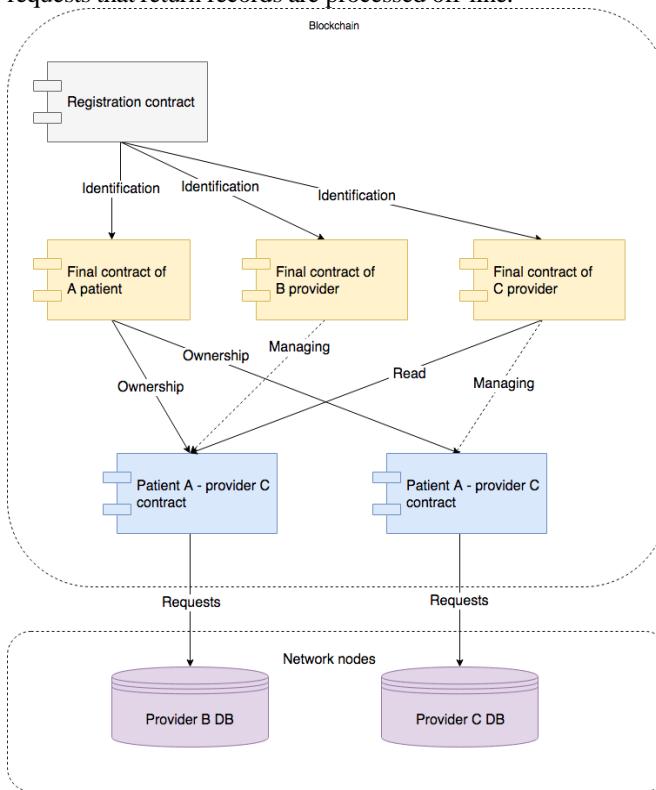


Fig. 1. Structure of patient-provider contract

Lets consider data processing in the system. A patient requests access to certain medical records by sending a request to the data provider that is part of the infrastructure outside the network of the system being developed. The data provider implements the interface of access to the local database of the patient node, which is governed by the rights that are stored in a flowchart. It runs a server that listens to requests that are cryptographically signed by the issuer from clients in the network. The cryptographic signature allows the gatekeeper to confirm the identity, and then checks blocking contracts to check if the requesting address is allowed to access the request. If the address is valid, it executes the request in the local database of the node and returns the result to the client.

It is assumed that many nodes of providers, especially those associated with service providers, already store data on networked servers with a high degree of security. The system also defines a modular protocol of interoperability that can interact with any application and user interface. Patient nodes also contain a local database, albeit more "light", which functions as a cache of patient data. The patient's node is a "light" node that can be executed on a PC or mobile phone.

Blockchain technologies introduce a number of confidentiality restrictions, some of which are alleviated by the use of the permitted read access structure and a private flowchart. The key issue is that even without a direct disclosure of the patient's name, the conclusion about who the particular patient is, can be made based on the metadata of one Ethereum address with several others. Even with a private

blockchain, there is a strong need to consider retrieving nodes that handle these sensitive metadata. One of the solutions is to require full permission from all the retrievers, and to require medical researchers working at retrieving centers only to provide secure systems.

The proposed solution to the problem of confidentiality is to use a system of "delegated contracts", where each provider creates separate Ethereum identifiers for each new relationship between patient providers. This means that instead of a single provider's address, from which the relations with particular patients can be easily obtained, the provider's identifier is distributed throughout the network. However, in order for the relationship to be safely established, the provider should not be able to add a new block containing this new address (since it would be easy to track each of these delegate addresses to the original). Therefore, when creating a new delegate account, the provider performs an arbitrary transaction with another verified provider by giving them details of the new delegate account that they can add as confirmed account information to the blockchain.

**Security.** The decentralized nature of the block-based systems gives the system the advantage of reliability both in authorization data support (stored in each node of the network) and in the repositories themselves (stored both by the patient and the corresponding provider node). With many organizations involved in the system, consensus mechanisms are also present to avoid separate points of failure. Since medical data and a global authorization log are distributed, there is no central goal for attacks or failures, and the network is intervention safe (since the modified node will conflict with other, unchanged nodes, thus making consensus impossible).

This system does not attempt to resolve security problems at the provider's database level (which must be duly managed by an IT service administrator), nor does it solve the security of the endpoint (a patient's compromised computer may potentially allow data theft).

Scalability is a constant concern in the Ethereum community and has not yet been resolved. One of the key issues is that any event stored at any time in a flowchart will appear in each subsequent block. Although this is also a feature of Bitcoin blockchain, since Ethereum provides both data storage and more complex operations, the effects of this growth are a big issue.

To integrate with the existing infrastructure of the electronic healthcare management system and records, it is necessary to design components of system nodes. Assume that many nodes and service providers in particular, are already trustingly managing databases with patient data stored on servers with network connections. The proposed system consists of four software components: a server unit library, an Ethereum client, a database gatekeeper and a medical records manager. They can be executed on servers, being united to create a consistent, distributed system. A prototype for implementing these components, which are integrated with the SQLite database and managed through a designed web-based user interface, is also offered. It should be noted that any implementation of the firewall and the user interface can participate in the system, using the module interaction protocol, defined through block diagrams.

Patient nodes in the proposed system contain the same basic components as suppliers. Their implementation can be done on a local computer or mobile phone. Their local database can be one of many light database implementations. Databases can function simply as a cache of patient medical data. Missing data can be obtained online at any time, following the final agreement of the center.

It is proposed to create server part library containing several utilities to facilitate the operation of the system. The library represents a connection to the blockchain and exports the API function. Management recording programs and their user interfaces, thus, can avoid interference with their direct work with the blockchain. One such obstacle is checking that each transaction sent is accepted with high trust from the network. The developed library automatically processes the indeterminacy when the transactions are retrieved and examines the cases when they are rejected. The backend library interacts with the Ethereum client to implement low-level formatting and analysis of the Ethereum protocol.

By using the blockchain registration contract, the patient's identifier first turns into the corresponding Ethereum address and the corresponding final contract is located. Then the provider downloads a new patient-provider relationship flowchart, indicating how they control the data belonging to the patient's Ethereum address. Then the provider's node creates a request for a link to this data and accordingly updates the patient-provider contract. Finally, the node sends a transaction that connects a new patient-provider contract with a patient's final contract, allowing the patient's nodes to later find it in a specific block.

Ethereum client implements the full functionality required for joining and participating in the Ethereum network. The client processes a wide range of tasks, such as: peer-to-peer network connection, encoding and sending transactions, and maintaining a verified local copy of the block template. The client has to be changed so that it enables the mapping of identity and addresses. Then the service is implemented to find the final contract of the node, by means of a register address search of the contract with the recorder. This service is constantly working within the client for monitoring real-time changes to the final contract. In the case of an update, the service signals that the medical data manager issues a user's message and, if necessary, synchronizes the local database. Modified Ethereum client constantly monitors its final contract. When a new block is retrieved from a newly-contracted patient-provider, the client makes a signal that results in a user's message. Then the user can confirm or refuse to communicate with the provider, accordingly updating the General Contract. If the message is accepted, then the implementation of the prototype automatically issues a request for new medical data. The client uses the information in a new patient-provider contract to locate the provider on the network and connect to its server's database gatekeeper.

The database gatekeeper implements an off-network access interface for the local node database, which is managed by the rights stored in blockchain. Gatekeeper launches a server that listens to requests from customers in the network. The request contains a request line/string, as well as a reference to the patient-provider contract, which requires permissions to run it. Request is cryptographically signed by the issuer,

which allows the gatekeeper to confirm its identity. Upon confirmation of the issuer's signature, the gatekeeper checks the contracts to determine whether access to the request is allowed on the requesting address. If the address is available, it executes the request in the local database of the node and returns the result to the client.

It should be noted that the created components in the same way support the receipt of patient data by third-party: the patient chooses the data to be sent and updates the corresponding patient-provider interaction contract with a third-party address and request line. If necessary, the patient's node may allow a third-party address using a registration contract. Then, the patient node connects an existing patient-provider contract with a third-party modifying provider. The third party is automatically notified of new permissions and can follow the link to find all the information needed. The gatekeeper of the provider's database will allow access to such a request, confirming that it was issued by the patient in a patient-provider shared user contract.

The medical data management system combines all of the software components mentioned above and the user interface. The program provides data from local SQLite databases (intended for interchange with other database software) to view and provides updates to users, as well as sharing and receiving data. The created user interface offers an intuitive, clear and informative design. The developed software system is conveniently accessible through a web interface built using JavaScript and AngularJS framework. Its compatibility with mobile devices is of particular note, since modern users expect easy and high-quality access from anywhere.

"Retrievers/(miners)" are encouraged to participate in the network and provide their computational resources to achieve a credible and gradual chain promotion. A model is proposed that embraces the medical community in the area of network management - the system developed involves health researchers and health stakeholders in their network. In turn, providers and patients give access to aggregated, anonymous medical data as a reward for retrieving/(mining). This idea is investigated in the developed prototype by introducing a special function in the patient-provider relationship contract. This requires providers to attach a request to any transaction they send by updating the patient-provider contract.

For example, this remuneration request can be arranged to return the average iron levels in the blood test done by the provider to all patients in the previous week. When the block containing the record-update operation is retrieved, the retrieving function automatically adds the block retriever as the request owner. The retriever can then collect it by simply sending a request for this reward to the provider's database gatekeeper. Since the unit is signed by the provider as part of the transaction, the remuneration request is safe from harmful changes. This "remuneration request" or retrieving/mining reward allows medical researchers to access data on medical treatment and health care outcomes at the community level. It is anticipated that future upgrades of the retrieving/mining model, where retrievers/miners can indicate the benefits for demographic groups and the peculiarities of the data which they seek, in order to provide accurate medicine and targeted research (while maintaining the confidentiality of patients).

## V. SOFTWARE SYSTEM IMPLEMENTATION

To implement the server part, the C# programming language and the .NET Framework were used together with connection of Ethereum platform modules. To implement the web client, it is used JavaScript programming language and AngularJS framework. The web site interface is adapted to mobile devices.

To begin, the user logs in to the system by sending photos of identification code and the first pages of the passport for verification. After verifying the data, the user creates a password, where the identification number acts as the password (Fig. 2).

By logging in with the ID number and password, the user is directed to the home page of the website (Fig. 3).

Here the history of all user's medical records, consolidated from various medical institutions, is presented.

Fig. 2. New patient registration page

| Visits and consultations |                           |              |                |                          |             |                         |
|--------------------------|---------------------------|--------------|----------------|--------------------------|-------------|-------------------------|
| Date                     | Hospital                  | Doctor       | Specialization | The purpose of the visit | Diagnosis   | Treatment               |
| 15.03.2017               | Ternopil city hospital №3 | Cayx O. B.   | Therapist      | Doctor's examination     | Healthy     | <button>REVIEW</button> |
| 15.03.2017               | Ternopil city hospital №1 | John A. B.   | Therapist      | Doctor's examination     | Healthy     | <button>REVIEW</button> |
| 15.03.2017               | Ternopil city hospital №3 | David A. R.  | Oculist        | Consultation             | -           | <button>REVIEW</button> |
| 15.03.2017               | Ternopil city hospital №3 | Mike R. V.   | Therapist      | Consultation             | Overfatigue | <button>REVIEW</button> |
| 15.09.2017               | Ternopil city hospital №1 | Pauler T. R. | Oculist        | Doctor's examination     | Overfatigue | <button>REVIEW</button> |

Fig. 3. Patient home page

When a medical institution wants to make a new record in history, notification is sent to the user, who can view the changes.

## VI. CONCLUSION

The prototype of a safe compatible system of medical data management based on Blockchain technology, which aggregates patient medical data from different providers, medical institutions' information systems, is developed. The system solves problems such as fragmented and slow access to patient medical data, incompatibility of medical information systems and improves the quality and quantity of data for medical research. Using intelligent Ethereum contracts to organize a content access system on separate sites for storage and provision of services, the authentication log determines access to medical records by providing patients with a comprehensive overview and data exchange. An innovative approach to integration with existing provider systems is demonstrated, identifying the priority of the open APIs and the transparency of the network structure.

## REFERENCES

- [1] Xia, Q.a, Sifah, E.B.b, Asamoah, K.O.b, Gao, J.c, Du, X.d, Guizani, M.e. Article "MeDShare: Trust-Less

Medical Data Sharing among Cloud Service Providers via Blockchain" IEEE, 22 July 2017, pp. 14757-14767.

- [2] Roehrs, A., da Costa, C.A. Author, da Rosa Righi, R. "OmniPHR: A distributed architecture model to integrate personal health records", *Journal of Biomedical Informatics*, July 2017, pp. 70-81.
- [3]. Zyskind, Guy, and Oz Nathan. "Decentralizing privacy: Using blockchain to protect personal data", *Security and Privacy Workshops (SPW)*, 2015, IEEE, pp. 180-184.
- [4] Nakamoto, Satoshi. "Bitcoin: A peer-to-peer electronic cash system", 2008.
- [5]. Kachmar V.O., "Medical Information systems – the state of development in Ukraine", *Ukrainian journal of telemedicine and medical telematics*, 2010, Vol.8, №1, p.12-17.
- [6]. Wood, Gavin. "Ethereum: A secure decentralised generalised transaction ledger", *Ethereum Project Yellow Paper*, 2014.
- [7]. Croman, Kyle, Christian Decker, Ittay Eyal, Adem Efe Gencer, Ari Juels, Ahmed Kosba, Andrew Miller, Prateek Saxena, Elaine Shi, and Emin Gün. "On scaling decentralized blockchains", *Financial Cryptography and Data Security. FC 2016*, pp. 106-125, 2016.

# Satisfiability Problems in Quasiary Program Logics

Mykola Nikitchenko, Stepan Shkilniak, Valentyn Tymofieiev

Department of Theory and Technology of Programming, Taras Shevchenko National University of Kyiv, UKRAINE,  
60 Volodymyrska Street, City of Kyiv, Ukraine, 01033, email: mykola.nikitchenko@gmail.com

**Abstract:** In the paper we present special program specification algebras and logics defined for classes of quasiary mappings. Informally speaking, such mappings are partial mappings defined over partial states (partial assignments) of variables. Conventional  $n$ -ary mappings can be considered as a special case of quasiary mappings. Such mappings better reflect properties of software systems. We describe methods of reducibility of the satisfiability problem in quasiary logics to the satisfiability problem in logics of  $n$ -ary mappings. The methods proposed can be useful for software verification.

**Keywords:**  $n$ -ary mapping, quasiary mapping, program algebra, specification logic.

## I. INTRODUCTION

Algebraic approach to software system specification has the following two characteristics: 1) the formalism of many-sorted algebras is used to model such systems; 2) special logics based on such many-sorted algebras are used to reason about system properties. In the literature various kinds of such algebras and logics are described (e.g. see [1, 2]).

In this paper we present special algebras and logics defined for classes of quasiary mappings. Informally speaking, such mappings are partial mappings defined over partial states (partial assignments) of variables. Conventional  $n$ -ary mappings can be considered as a special case of quasiary mappings. Quasiary mappings better reflect properties of software systems therefore construction and investigation of algebras and logics of quasiary mapping is an important challenge.

Proposed constructions are based on a composition-nominative approach [3]. Principles of the approach (*development of program notions from abstract to concrete, priority of semantics, compositionality of programs, and nominativity of program data*) specify program models as *composition-nominative systems* which consist of *composition, description, and denotation* systems. A *composition* system defines semantic aspects of programs, a *description* system defines program descriptions (syntactic aspects), and a *denotation* system specifies meanings (referents) of descriptions. We consider semantics of programs as partial functions over a class of data processed by programs; compositions are  $n$ -ary operations over functions. Thus, a composition system can be specified by two algebras: *data algebra* and *function algebra*. Function algebra is the main *semantic notion* in program formalization. Terms of this algebra define *syntax* of programs (description system), and ordinary procedure of term interpretation gives a *denotation* system.

The constructed program models form a base for developing special program logics called *composition-*

*nominative logics* (CNL).

In this paper we continue our work on studying CNL [4–6] focusing on quasiary specification algebras and logics. The main questions under discussion concern satisfiability problems and their reduction to satisfiability problems in logics of  $n$ -ary mappings.

## II. QUASIARY MAPPINGS

Quasiary mappings can be met in different branches of mathematics, logics, and computer science. Informally speaking, such mappings appear when we use variables (names) to construct mapping arguments. Here we consider only usage of quasiary mappings in logic semantics and formal models of programs.

The notion of quasiary predicate and function can be easily understood when we analyze Tarski's definition of first-order language semantics. This semantics is based on the notion of interpretation which consists of two parts: 1) interpretation of predicate and function symbols in some structure, and 2) interpretation of individual variables in the domain of this structure. The latter are usually called *variable assignments (or valuations)* and can be represented by total mappings from a set of individual variables (names)  $V$  into some set of basic values  $A$ . The class of such total mappings will be denoted  $V \xrightarrow{\cdot} A$  or  $A^V$ , and called *total nominative sets*. Thus, Tarski's semantics interprets predicate and function symbols as total quasiary predicates and functions defined on the class  $A^V$  of total nominative sets. In applications like model checking, program verification, automated theorem proving, etc., partial assignments (nominative sets) are often used instead of total assignments. The class of such partial mappings will be denoted  $V \xrightarrow{P} A$  or  $V^A$ , and called *partial nominative sets* (partial data); the term 'partial' is often omitted. Predicates and functions over nominative sets are called *quasiary*. This means that formulas and terms can be interpreted as quasiary predicates and functions respectively.

Quasiary mappings also appear in a natural way in denotational semantics of programs. In this semantics program states are represented as nominative sets, Boolean expressions as *quasiary predicates* of the type  $Pr_A^V =^V A$   $\xrightarrow{P} \text{Bool}$ , arithmetical expressions as *quasiary functions* of the type  $Fn_A^V =^V A \xrightarrow{P} A$ , and program statements as *bi-quasiary functions (program functions)* of the type  $PF_A^V =^V A$   $\xrightarrow{P} V^A$ . Semantics of structured statements is defined by the following compositions with conventional meaning: *assignment composition AS<sup>x</sup>* ( $x$  is a parameter from  $V$ ), *composition of sequential execution*  $\bullet$ , *conditional*

*composition IF, loop composition WH.* For structural expressions we additionally use unary denomination composition ' $x$ ' and various superpositions.

Thus, we obtain a program algebra with three carriers: quasiary predicates, quasiary functions, and bi-quasiary functions (program functions). Such algebras can be called *algorithmic algebras*.

To extend such algebras to program specification algebras we add *quantifiers* and *prediction composition* ' $\cdot$ '. Prediction composition is simply a functional composition of a program function and a predicate. This composition is strong enough to represent Hoare assertions, and therefore, specification algebras with these compositions are rather expressive [7].

In the rest of the paper we consider *program specification algebras* and *logics of partial quasiary mappings*.

To emphasize a mapping's *partiality/totality* we write the sign  $\xrightarrow{p}$  for partial mappings and the sign  $\xrightarrow{t}$  for total mappings. Given a partial mapping  $\mu, \mu': D \xrightarrow{p} D'$ ,  $d, d' \in D$  we write:

- $\mu(d) \downarrow (\mu(d) \uparrow)$  to denote that  $\mu$  is defined (undefined) on  $d$ ;
- $\mu(d) \downarrow = d'$  to denote that  $\mu$  is defined on  $d$  with a value  $d'$ ;
- $\mu(d) \cong \mu'(d')$  to denote the strong equality.

We omit proofs and some details of complicated definitions.

### III. QUASIARY SPECIFICATION ALGEBRAS

We use the following set of composition symbols parameterized by  $V$ :

$$Cs^V = \{\vee, \neg, \exists x, 'x, S_P^{v_1, \dots, v_n}, S_F^{v_1, \dots, v_n}, =, S_G^{v_1, \dots, v_n}, AS^x, \bullet, IF, WH, \cdot\}$$

Formal definitions of compositions can be found in [4, 7, 8]. Compositions  $S_P^{v_1, \dots, v_n}$ ,  $S_F^{v_1, \dots, v_n}$ , and  $S_G^{v_1, \dots, v_n}$  are called superpositions and represent substitutions of quasiary functions into a predicate, function, and program function respectively. We also denote such compositions as  $S_P^{\bar{v}}$ ,  $S_F^{\bar{v}}$ ,  $S_G^{\bar{v}}$ .

A tuple  $\mathfrak{A}_A^V(Cs^V) = \langle Pr_A^V, Fn_A^V, PF_A^V; Cs^V \rangle$  is called *quasiary specification algebra* (QSA) over  $V$  and  $A$ .

Variables used as composition parameters can be classified as *essential* in the sense that they can affect the result of composition application evaluation and as *updatable* in the sense that the values of these variables can change during evaluation. Variable  $x$  is essential for denomination composition ' $x$ ', variable  $x$  is updatable for  $\exists x$  and  $AS^x$ , variables  $v_1, \dots, v_n$  are updatable for  $S_P^{v_1, \dots, v_n}$ ,  $S_F^{v_1, \dots, v_n}$ , and  $S_G^{v_1, \dots, v_n}$ .

Now we describe the main properties of superpositions.

**Lemma 1** (superposition folding). Let  $\bar{y} = y_1, \dots, y_n$ ,  $\bar{f}' = f_1', \dots, f_n'$ ;  $\bar{x} = x_1, \dots, x_k$ ,  $\bar{f} = f_1, \dots, f_k$ ,  $\bar{h}' = h_1', \dots, h_k'$ ,  $\bar{v} = v_1, \dots, v_m$ ,  $\bar{h} = h_1, \dots, h_m$ ,  $\{y_1, \dots, y_n\} \cap \{v_1, \dots, v_m\} = \emptyset$ . Then the following properties hold in  $\mathfrak{A}_A^V(Cs^V)$ :

$$SS_P. \quad S_P^{\bar{y}, \bar{x}}(S_P^{\bar{x}, \bar{v}}(p, \bar{f}, \bar{h}), \bar{f}', \bar{h}') = S_P^{\bar{y}, \bar{x}, \bar{v}}(p, \bar{\sigma}),$$

$$SS_F. \quad S_F^{\bar{y}, \bar{x}}(S_F^{\bar{x}, \bar{v}}(f, \bar{f}, \bar{h}), \bar{f}', \bar{h}') = S_F^{\bar{y}, \bar{x}, \bar{v}}(f, \bar{\sigma}),$$

$$SS_G. \quad S_G^{\bar{y}, \bar{x}}(S_G^{\bar{x}, \bar{v}}(g, \bar{f}, \bar{h}), \bar{f}', \bar{h}') = S_G^{\bar{y}, \bar{x}, \bar{v}}(g, \bar{\sigma}),$$

where

$$\bar{\sigma} = (\bar{f}', S_F^{\bar{y}, \bar{x}}(f_1, \bar{f}', \bar{h}'), \dots, S_F^{\bar{y}, \bar{x}}(f_k, \bar{f}', \bar{h}'), \\ S_F^{\bar{y}, \bar{x}}(h_1, \bar{f}', \bar{h}'), \dots, S_F^{\bar{y}, \bar{x}}(h_m, \bar{f}', \bar{h}')).$$

**Lemma 2** (distributivity of superposition). The following properties hold in  $\mathfrak{A}_A^V(Cs^V)$ :

$$S \vee. \quad S_P^{\bar{v}}(p \vee q, \bar{f}) = S_P^{\bar{v}}(p, \bar{f}) \vee S_P^{\bar{v}}(q, \bar{f}),$$

$$S \neg. \quad S_P^{\bar{v}}(\neg p, \bar{f}) = \neg S_P^{\bar{v}}(p, \bar{f}),$$

$$S =. \quad S_P^{\bar{v}}(h_1 = h_2, \bar{f}) = (S_F^{\bar{v}}(h_1, \bar{f}) = S_F^{\bar{v}}(h_2, \bar{f})),$$

$$S \exists. \quad S_P^{\bar{v}}(\exists x p, \bar{f}) = \exists u S_P^{\bar{v}}(S_P^x(p, 'u), \bar{f}), \quad u \not\equiv x, \quad u \notin \bar{v}, \quad u \text{ is unessential for } p \text{ and } \bar{f},$$

(here " $u$  is unessential for  $p$  and  $\bar{f}$ " means that  $p(d) \equiv p(d')$  and  $\bar{f}(d) \equiv \bar{f}(d')$  for any  $d$  and  $d'$  such that  $d \parallel_{\{u\}} d' \parallel_{\{u\}}$ ).

Superposition compositions are not distributive with respect to *WH*, therefore we simplify superpositions into program functions using the identity program function *id*.

**Lemma 3** (superpositions with program functions). The following properties hold in  $\mathfrak{A}_A^V(Cs^V)$ :

$$SG. \quad S_G^{\bar{v}}(g, \bar{f}) = S_G^{\bar{v}}(id, \bar{f}) \bullet g - \text{superposition into program function,}$$

$$SP. \quad S_P^{\bar{v}}(g \cdot p, \bar{f}) = S_P^{\bar{v}}(g, \bar{f}) \cdot p - \text{superposition with prediction composition.}$$

**Lemma 4** (superposition simplification). The following properties hold in  $\mathfrak{A}_A^V(Cs^V)$ :

$$SE. \quad S_P(p) = p, S_F(f) = f, S_G(g) = g - \text{superpositions with empty parameter list,}$$

$$SiD. \quad S_F^{\bar{v}}('x, \bar{f}) = 'x \quad \text{if} \quad x \notin \bar{v}, S_F^{x, \bar{v}}('x, f, \bar{h}) = f - \text{superpositions into denomination functions,}$$

$$SwD. \quad S_P^{x, \bar{v}}(p, 'x, \bar{f}) = S_P^{\bar{v}}(p, \bar{f}), \quad S_F^{x, \bar{v}}(h, 'x, \bar{f}) = S_F^{\bar{v}}(h, \bar{f})$$

- superpositions with denomination function,

$$ST. \quad S_P^{\bar{v}, x, y, \bar{z}}(\varphi, \bar{f}, h, h', \bar{f}') = S_P^{\bar{v}, y, x, \bar{z}}(\varphi, \bar{f}, h', h, \bar{f}'),$$

$$S_F^{\bar{v}, x, y, \bar{z}}(f, \bar{f}, h, h', \bar{f}') = S_F^{\bar{v}, y, x, \bar{z}}(f, \bar{f}, h', h, \bar{f}'),$$

$$S_G^{\bar{v}, x, y, \bar{z}}(g, \bar{f}, h, h', \bar{f}') = S_G^{\bar{v}, y, x, \bar{z}}(g, \bar{f}, h', h, \bar{f}') - \text{transposition of parameters.}$$

These properties will be used to construct superpositional normal forms for language expressions.

Now we study relations between QSA  $\mathfrak{A}_A^V(Cs^V)$  and QSA  $\mathfrak{A}_A^{V'}(Cs^{V'})$  induced by the following two relations between sets of their names:

1) there is a renomination bijection  $\beta: V \xrightarrow{t} V'$ ,

2)  $V'$  is an extension of  $V$  ( $V \subseteq V'$ ).

In the first case  $\beta$  induces in a natural way new mappings  $\beta_P: Pr_A^V \xrightarrow{t} Pr_A^{V'}$ ,  $\beta_F: Fn_A^V \xrightarrow{t} Fn_A^{V'}$ , and  $\beta_G: PF_A^V \xrightarrow{t} PF_A^{V'}$  with the following properties.

**Theorem 1.** Mappings  $\beta_C$ ,  $\beta_P$ ,  $\beta_F$ , and  $\beta_G$  define an isomorphism of QSA  $\mathfrak{A}_A^V(Cs^V)$  and QSA  $\mathfrak{A}_A^{V'}(Cs^{V'})$ .

**Theorem 2.** Let  $V \subseteq V'$ . Then inclusion mapping induces (ignoring variables from  $U = V' \setminus V$ ) an injective homomorphism of  $\mathfrak{A}_A^V(Cs^V)$  into  $\mathfrak{A}_A^{V'}(Cs^{V'})$ .

Now we can study an algebra with mappings over total data that ‘‘mimic’’ mappings over partial data. A special element  $\varepsilon$  ( $\varepsilon \notin A$ ) will represent a case when a value of a variable or a function is undefined. Let  $A_\varepsilon = A \cup \{\varepsilon\}$  and  $A_\varepsilon^V = V \xrightarrow{t} A \cup \{\varepsilon\}$ . We construct a QSA  $\mathfrak{A}_{A,\varepsilon}^V(Cs_\varepsilon^V)$  with total data that ‘‘mimics’’ QSA  $\mathfrak{A}_A^V(Cs^V)$ . Carriers of the new algebra are classes  $Pr_{A,\varepsilon}^V = A_\varepsilon^V \xrightarrow{p} \text{Bool}$ ,  $FnT_{A,\varepsilon}^V = A_\varepsilon^V \xrightarrow{t} A_\varepsilon$ , and  $PF_{A,\varepsilon}^V = A_\varepsilon^V \xrightarrow{p} A_\varepsilon^V$ .

Then we define mapping  $\varepsilon_D^+ : V \xrightarrow{t} A_\varepsilon^V$  that ‘‘add  $\varepsilon$  into a nominative set. This mapping induces mappings  $\varepsilon_P^+$ ,  $\varepsilon_F^+$ , and  $\varepsilon_G^+$  relating corresponding carriers.

**Theorem 3.** Mappings  $\varepsilon_P^+$ ,  $\varepsilon_F^+$ , and  $\varepsilon_G^+$  define an isomorphism of  $\mathfrak{A}_A^V(Cs^V)$  and  $\mathfrak{A}_{A,\varepsilon}^V(Cs_\varepsilon^V)$ .

We treat  $n$ -ary operations as a special case of quasianary mappings with the set of variables  $N = \{1, \dots, n\}$  and total data. In this case a total nominative set  $[1 \mapsto a_1, \dots, n \mapsto a_n]$  is represented by a tuple  $(a_1, \dots, a_n)$ . Thus, all algebra mappings are defined on a Cartesian product  $A^n$ . Compositions from  $Cs^N$  can be treated as compositions over  $n$ -ary mappings.

Here we do not redefine compositions in this style assuming that it is a simple task. The term ‘‘unified’’ means that all mappings have the same arity.

Obtained algebra is called a unified  $n$ -ary specification algebra (NSA) and is denoted  $\mathfrak{A}_A^N(Cs^N)$ . The following proposition is practically an immediate consequence of Theorems 1–3.

**Theorem 4.** Let  $V = \{v_1, \dots, v_n\}$ ,  $N = \{1, \dots, n\}$  ( $n \geq 1$ ),  $\beta : V \xrightarrow{t} N$  be a bijection,  $\mathfrak{A}_A^V(Cs^V)$  be QSA and  $\mathfrak{A}_{A,\varepsilon}^N(Cs^N)$  be NSA ( $\varepsilon \notin A$ ). Then mappings  $\beta_C$ ,  $\beta_P \circ \varepsilon_P^+$ ,  $\beta_F \circ \varepsilon_F^+$ , and  $\beta_G \circ \varepsilon_G^+$  define an isomorphism of QSA  $\mathfrak{A}_A^V(Cs^V)$  and NSA  $\mathfrak{A}_{A,\varepsilon}^N(Cs^N)$ .

#### IV. QUASIARY SPECIFICATION LOGIC

To define a *quasiary specification logic*, denoted  $\mathcal{L}^Q$ , we have to specify its semantic, syntactic, and interpretational components [4, 8].

Semantic components of  $\mathcal{L}^Q$  is based on the class of quasiary specification algebras  $\mathfrak{A}_A^V(Cs^V)$  for different  $A$ .

A syntactic component specifies the language of  $\mathcal{L}^Q$  constructed over signature  $\Sigma_Q^V = (Cs^V, Ps, Fs, Pgs)$  where  $Ps$ ,  $Fs$ , and  $Pgs$  are respectively the sets of predicate symbols, ordinary function symbols, and program function symbols. For simplicity, we use the same notation for symbols of compositions and compositions themselves.

For a given signature  $\Sigma_Q^V$  the set of formulas  $Fr(\Sigma_Q^V)$ , the

set of terms  $Tr(\Sigma_Q^V)$ , and the set of programs  $Pg(\Sigma_Q^V)$  are defined by induction in a traditional way.

Interpretational component is defined in the following way. Given  $\Sigma_Q^V = (Cs^V, Ps, Fs, Pgs)$  and a set  $A$  we can define a QSA  $\mathfrak{A}_A^V(Cs^V) = \langle Pr_A^V, Fn_A^V, PF_A^V; Cs^V \rangle$ . Composition symbols have fixed interpretation, but we additionally need interpretations  $I^{Ps} : Ps \xrightarrow{t} Pr_A^V$ ,  $I^{Fs} : Fs \xrightarrow{t} Fn_A^V$ , and  $I^{Pgs} : Pgs \xrightarrow{t} PF_A^V$  of predicate, function, and program function symbols respectively. A tuple  $J = (\Sigma_Q^V, A, I^{Ps}, I^{Fs}, I^{Pgs})$  is called an  $\mathcal{L}^Q$ -interpretation. Usually the prefix  $\mathcal{L}^Q$  is omitted. Given an interpretation  $J$  we denote meanings in  $J$  of a formula  $\Phi$ , a term  $t$ , and a program  $\pi$  respectively  $\Phi_J$ ,  $t_J$ , and  $\pi_J$ .

$\mathcal{L}^Q$ -formula  $\Phi$  is *satisfiable in an interpretation J* if there exists an element  $d$  such that  $\Phi_J(d) \downarrow = T$ . This is denoted  $\mathcal{L}^Q, J \models \Phi$ . Formula  $\Phi$  is *satisfiable in the logic  $\mathcal{L}^Q$*  ( $\mathcal{L}^Q \models \Phi$ ), if there exists an interpretation  $J$  such that  $\mathcal{L}^Q, J \models \Phi$ . Formulas  $\Phi$  and  $\Psi$  are *equisatisfiable*, if they are both satisfiable or both unsatisfiable.

$\mathcal{L}^Q$ -formula  $\Phi$  is called *valid in an interpretation J* if there is no  $d$  such that  $\Phi_J(d) \downarrow = F$ . This is denoted  $\mathcal{L}^Q, J \vdash \Phi$ , which means that  $\Phi$  is not refutable in  $J$ . A formula  $\Phi$  is called *valid in  $\mathcal{L}^Q$*  if  $\mathcal{L}^Q, J \vdash \Phi$  for any interpretation  $J$ . We shall denote this  $\mathcal{L}^Q \models \Phi$ , or just  $\models \Phi$  if the logic in hand is understood from the context.

$\mathcal{L}^Q$ -formulas  $\Phi$  and  $\Psi$  are *equivalent*, if for every  $J$  predicates  $\Phi_J$  and  $\Psi_J$  are identical. Such notion of equivalence can be also defined for terms and programs.

Validity and satisfiability problems for QSL are related in the following way:  $\Phi$  is valid in  $J$  if and only if  $\neg\Phi$  is unsatisfiable in  $J$ .

Let  $N = \{1, \dots, n\}$ . We treat a unified  $n$ -ary specification logic  $\mathcal{L}^N$  as a quasiary specification logic with a signature  $\Sigma_N^n = (Cs_n^{\{1, \dots, n\}}, Ps, Fs, Pgs)$  constructed over total nominative sets. This logic is semantically based on unified  $n$ -ary specification algebras.

Now we will study a problem how to relate  $\mathcal{L}^Q$  and  $\mathcal{L}^N$  with respect to the satisfiability problem, namely, given  $\mathcal{L}^Q$ -formula  $\Phi$  construct  $\mathcal{L}^N$ -formula  $\Phi^n$  such that  $\Phi$  and  $\Phi^n$  will be equisatisfiable. We do this in several steps:

- introducing a logic  $\mathcal{L}^{QU}$  with unessential variables,
- constructing a superpositional normal form  $\Phi^s$  of  $\Phi$  in  $\mathcal{L}^{QU}$ ,
- introducing a logic  $\mathcal{L}^{QUR}$  with finitely restricted sets of updatable variables,
- constructing a unified superpositional normal form  $\Phi^u$  of  $\Phi^s$  in  $\mathcal{L}^{QUR}$ ,
- constructing from  $\Phi^s$  a formula  $\Phi^t$  of logic  $\mathcal{L}_T^{QUR}$  with total data,
- translating  $\Phi^t$  into  $\mathcal{L}^N$ -formula  $\Phi^n$ ,

— proving equisatisfiability of  $\Phi$  and  $\Phi^n$ .

Logic  $\mathcal{L}^Q$  being a rather powerful logic still is not expressible enough to represent various important transformations. Therefore we introduce as its extension a logic with unessential variables denoted  $\mathcal{L}^{QU}$ . Here  $U$  is an infinite set of variables such that  $V \cap U = \emptyset$ . Unessential variables do not affect the meaning of formulas (terms, programs) [4,8]. An additional requirement is that unessential variables are not updatable by programs. The signature of  $\mathcal{L}^{QU}$  is  $\Sigma_Q^{V \cup U} = (Cs^{V \cup U}, Ps, Fs, Pgs)$ .

The following statement is a consequence of Theorem 2.

**Lemma 5** ( $\mathcal{L}^{QU}$  is a model-theoretic conservative extension of  $\mathcal{L}^Q$ ). Let  $\mathcal{L}^{QU}$ -interpretation  $J^U$  be an unessential extension of  $\mathcal{L}^Q$ -interpretation  $J$ ,  $\Phi \in Fr(\Sigma_Q^V)$ ,  $t \in Tr(\Sigma_Q^V)$ ,  $\pi \in Pg(\Sigma_Q^V)$ . Then

$$i_P^U(\Phi_J) = \Phi_{J^U}, i_F^U(t_J) = t_{J^U}, \text{ and } i_G^U(\pi_J) = \pi_{J^U}.$$

Introduction of  $\mathcal{L}^{QU}$  permits to formulate transformations rules based on properties presented in Lemmas 1–4.

$\mathcal{L}^{QU}$ -formula  $\Phi$  is said to be in *superpositional normal form*, if the following conditions hold:

SP. For each subformula  $S_P^{\bar{v}}(\Psi, \bar{t})$  of formula  $\Phi$  we have that  $\Psi \in Ps$ ;

SF. For each subformula of the form  $S_F^{\bar{v}}(t, \bar{t})$  we have that  $t \in Fs$ ;

SG. For each subformula of the form  $S_G^{\bar{v}}(\pi, \bar{t})$  we have that  $\pi = id$ .

**Lemma 6.** Let  $\Phi \in Fr(\Sigma_Q^{V \cup U})$ . Then, using transformation specified by Lemmas 1–4, a superpositional normal form  $\Phi^s$  of  $\Phi$  can be constructed such that  $\Phi^s \approx \Phi$ .

In a similar way we can define transformations that first lead to a formula  $\Phi'$  of logic  $\mathcal{L}_T^{QUR}$  with total data and then to formula  $\Phi^n$  of logic  $\mathcal{L}^N$ .

## V. REDUCTION OF THE SATISFIABILITY PROBLEM

Combining all obtained results, we can prove the following main theorem that states reducibility of the satisfiability problem in quasiary specification logics with finitely restricted sets of updatable variables to the satisfiability problem in  $n$ -ary specification logics.

**Theorem 5.** Let  $\Phi$  be a  $\mathcal{L}^{QUR}$ -formula and  $\Phi^n$  be a  $\mathcal{L}^N$ -formula obtained by the above-described transformations. Then  $\Phi$  and  $\Phi^n$  are equisatisfiable.

Results of such kind permit to use existing satisfiability checkers for classical predicate and program logics, based on  $n$ -ary mappings, to check satisfiability of formulas for quasiary logics.

## VI. CONCLUSION

In this paper, we have developed special program specification algebras and logics defined for classes of quasiary mappings. These algebras and logics reflect such

features of software systems as partiality of data, partiality and unrestricted arity of predicate and functions, sensitivity to unassigned variables. For the constructed logics some laws of classical logic fail. We have studied relations of quasiary logics to logics of  $n$ -ary mapping. Obtained results demonstrate that logics of quasiary mappings are more powerful and expressive than logics based on  $n$ -ary mappings. We have developed methods of reduction of the satisfiability problems in quasiary logics to the satisfiability problems for logics based on  $n$ -ary mappings. Such methods can be useful for construction and investigation of logics for program reasoning.

Future work on the topic will include construction of calculi for important fragments of the considered logics. Also, a prototype of software systems for theorem proving in quasiary specification logics should be developed. First steps in this direction are made in [9, 10].

## REFERENCES

- [1] Handbook of Logic in Computer Science, S. Abramsky, Dov M. Gabbay, and T. S. E. Maibaum (eds.), in 5 volumes, Oxford Univ. Press, Oxford, 1993–2001.
- [2] D. Sannella, A. Tarlecki, “Foundations of Algebraic Specification and Formal Software Development”, Springer, 2012.
- [3] N.(M.) Nikitchenko, “A Composition Nominative Approach to Program Semantics”. *Technical Report IT-TR 1998-020*, Technical University of Denmark, 103 p., 1998.
- [4] M. Nikitchenko, V. Tymofieiev, “Satisfiability in Composition-Nominative Logics”, *Central European Journal of Computer Science*, vol. 2, issue 3, 2012, pp. 194–213.
- [5] M. Nikitchenko, S. Shkilniak, “Applied Logic”, Publishing house of Taras Shevchenko National University of Kyiv, Kyiv, 2013 (in Ukrainian), 278 p.
- [6] M. Nikitchenko, V. Tymofieiev, “Composition-Nominative Logics in Rigorous Development of Software Systems”, *LNBIP*, vol. 137, pp. 140–151. Springer, Heidelberg, 2013.
- [7] A. Kryvolap, M. Nikitchenko, W. Schreiner, “Extending Floyd-Hoare logic for partial pre- and postconditions”, *CCIS*, vol. 412, pp. 355–378, Springer, Heidelberg, 2013.
- [8] M. Nikitchenko, S. Shkilniak, “Algebras and logics of partial quasiary predicates”, *Algebra and Discrete Mathematics*, vol. 23, number 2, 2017, pp. 263–278.
- [9] I. Ivanov, M. Nikitchenko, A. Kryvolap, A. Kornilowicz, “Simple-Named Complex-Valued Nominative Data – Definition and Basic Operations”, *Formalized Mathematics*, 25(3), pp. 205–216, 2017.
- [10] A. Kornilowicz, A. Kryvolap, M. Nikitchenko, I. Ivanov, “Formalization of the nominative algorithmic algebra in Mizar”, *Advances in Intelligent Systems and Computing*, vol. 656, pp. 176–186, Springer, 2017.

# Software System for Formation the Composition of Academic Groups (Subgroups) Based on the Diffusion-Like Model

Natalia Porplytsya, Serhii Dubovyi

Department of Computer Science, Ternopil National Economic University, UKRAINE, Ternopil, 8 Chehova str.,  
email: [ocheretnyuk.n@gmail.com](mailto:ocheretnyuk.n@gmail.com), [dubovij.sergij@gmail.com](mailto:dubovij.sergij@gmail.com)

**Abstract:** In this paper, the method of forming a composition of academic groups (subgroups) in an educational institution was developed. The method based on a diffusion-like model of the dissemination of knowledge potential. The application of this method ensures the formation of academic groups (subgroups) in such a way as to maximize the coefficient of distribution of knowledge potential within each of the formed groups (subgroups). Software system, which implements the specified method, is also considered.

**Keywords:** diffusion-like model, knowledge potential, academic group.

## I. ACTUALITY OF THE PROBLEM

Student academic group is a kind of small group, with its stages of transformation into a collective, with parameters of development and criteria of formation. Constant studying of the level of development and knowledge of each student and the team of the academic group enables to effectively develop the educational process in a higher educational institution, taking into account the changes that the student's team undergoes in general and each member in particular, to correct the content and methodology of this process [1-5].

The initial formation of academic groups is based on the level of knowledge of each entrant, which usually reflects as the average rating point for the implementation of certification works on selected external subjects of external independent testing (EIT).

External independent testing (EIT, formerly also External testing, ET) - entrance exams for higher education in Ukraine. The complex of organizational procedures (first of all - testing) aimed at determining the level of educational achievements of graduates from secondary schools when they enter higher education institutions [6]. At the same time, the division of students into groups (subgroups) in educational institutions, usually carried out on the basis of the average rating point without taking into account its detail that is, the compliance of competing subjects of EIT chosen specialty. An applicant may receive a high score from the "Ukrainian language and literature" and at the same time and a low from the "Mathematics" at the entrance to the technical specialty, where his level of knowledge plays an important role in the field of mathematics. In addition, in other cases, the formation of academic groups from the list enrolled to study entrants can be carried out in alphabetical order, without considering at the same time the level of knowledge of applicants, which does not provide in the future the maximum effectiveness of the educational process. Further, during the studing process, formed training groups for

laboratory (practical) classes on individual disciplines can be divided into subgroups. In this case, such a division is usually carried out in alphabetical order. Such an approach is completely unreasonable, since it does not take into account the assessments received by students for related disciplines in accordance with the structural-logical scheme specialty.

Therefore, it is important to develop a method for forming a composition of academic groups (subgroups) in an educational institution, in such a way as to ensure the maximum effectiveness of the educational process. To construct this method expedient to use a diffusion-like model of the propagation of the knowledge potential, which simulates the transfer of knowledge between students by analogy with the crystallization process of the solid body from the melt at the outlet from the heat [7, 8]. In addition, it's need to develop the software module for the implementation of this method.

## II. MODELING OF THE COMPOSITION OF ACADEMIC GROUPS AND SUBGROUPS BY APPLICATION OF THE KNOWLEDGE POTENTIAL DISTRIBUTION COEFFICIENT

During the formation of the composition of academic groups, we will take into account the calculated coefficient of distribution of the knowledge potential calculated for them [8], that is, the higher this value is the better. It is important to note that the possible maximum value of this coefficient is unknown in advance. In addition, the distribution of compositions to academic groups or subgroups is carried out in such a way as to ensure the highest and virtually equivalent values of these coefficients for all formed academic units. Then the task of finding such a distribution of student compositions can be equationted in the form of a multivariate discrete optimization problem.

At the same time, we will examine all possible variants of the composition of the academic groups, which in turn shows that this task is extremely complex and belongs to the NP-complex class. For each combination, the value of the coefficient of distribution of the knowledge potential of the group will be compared, which will be based on the level of knowledge of each student, taking into account the individual characteristics of each discipline and specialty.

So, determination of the coefficient of distribution of the known potential will be calculated  $C_n^k$  times, where  $C_n^k$  - number of combinations from n to k; k - number of students in the academic group (subgroup); n - total number of

students;  $m$  – required number of groups (subgroups). It will be received  $C_n^k$  sets of coefficients of distribution knowledgeable potential  $\vec{b}_l = b_1 \dots b_m, i = 1 \dots k$  which we will store, for further determination of the "best" according to the following criteria:

$$\begin{cases} b_1 \rightarrow \max, b_1 \in \vec{b}_l; \\ \dots \\ b_m \rightarrow \max, b_m \in \vec{b}_l; \\ |b_1 - b_2| \rightarrow 0, b_1, b_2 \in \vec{b}_l; \\ \dots \\ |b_1 - b_{m-1}| \rightarrow 0, b_{m-1}, b_m \in \vec{b}_l; \end{cases} \quad (1)$$

For example, when dividing students into two subgroups will be implemented  $C_n^k / 2$  comparisons of the distribution coefficients of the knowledge potential of the formed subgroups As can be seen from expression (1), as a result of the test, an option is chosen for which the difference in the knowledge potential among the groups formed will be the smallest, and the value of the knowledge potential will be greatest. One of the options for introducing generalized potential  $K_j$ -group  $\varphi_{j,m}$ . There is a representation of it in the form of some function from  $\varphi_{j,k,m}$ , in particular, in the form of a generalized arithmetic mean:

$$\varphi_{j,m} = \frac{1}{k_j} \sum_{k=1}^{k_j} \alpha_{j,k} \varphi_{j,k,m} \quad (2)$$

where  $\alpha_{j,k}$  – some weight factor.

In the initial iteration of the use of the method that is, the formation of a composition of academic groups of students of the first year will determine the knowledge potential of each student based on the results of the EIT. In this case, each item will be given a coefficient of importance that corresponds to the chosen specialty. Thus, the knowledge potential of a particular student will look like:

$$\varphi_{j,k,m} = \sum_{i=1}^t z_i p_i \quad (3)$$

where  $z_i$  – coefficient of importance of an object,  $p_i$  – subject score,  $t$  – number of subjects. The sum of the coefficients of importance  $\sum_{i=1}^t z_i = 1$ .

To form a composition of subgroups for the subject, you should introduce the notion of the source of knowledge, which will be the lecturer who will conduct practical classes. Depending on the qualification, each lecturer will be assigned the appropriate efficiency factor, which will characterize him as a source of knowledge. Table one shows the adequacy of the efficiency coefficients for the qualifications obtained.

TABLE 1. CONFORMITY OF THE POINTS FOR THE CALCULATION OF THE QUALIFICATIONS

| Obtained qualification | Level of qualification | Efficiency coefficient |
|------------------------|------------------------|------------------------|
| Lecturer trainee       | 1                      | 60                     |
| Lecturer               | 2                      | 65                     |
| Lecturer, Ph.D.        | 3                      | 80                     |
| Senior Lecturer, Ph.D. | 4                      | 90                     |
| Docent, Ph.D.          | 5                      | 100                    |

The formation of subgroups of students from groups in the first semester of the first year will also be based on the results of the EIT. Thus, the equation for determining the coefficient of distribution of the knowledge potential of the formed subgroup  $A_{j,g,m+1}$  will look like:

$$A_{j,g,m+1} = \sum_{j=0}^j \varphi_{j,k,m+1} - \varphi_{j,k,m} \text{ where} \\ \varphi_{j,k,m+1} - \varphi_{j,k,m} = f_{j,k,m} + \\ + D_{j,k,m} \sum_{1 \leq k < k < k \leq k_j} \sigma_{k,k,\bar{k}} (\varphi_{j,k,m} - 2\varphi_{j,k,m} + \varphi_{j,k,m}), \quad (4)$$

$f_{j,k,m}$  - source of knowledge,  $D_{j,k,m}$  - coefficient which characterizing the ability  $k$ -agent  $j$ -group redistribute information (knowledge) at the time  $m$  (analog of diffusion coefficient)[7].

It should be noted that for the use of equation (4), the student's marks with the EIT, which are calculated in the 200-point system, should be converted into 100-point system. It is known that the minimum passing point of the EIT is 100 points, while the minimum score required for passing the discipline at the university is 60 points. We will make the appropriate proportion for transfer of points from the 200-point system to 100 points, which is represented by the equation:

$$x = \frac{60 * c}{100}, \quad (5)$$

where  $c$  – EIT score for a subject.

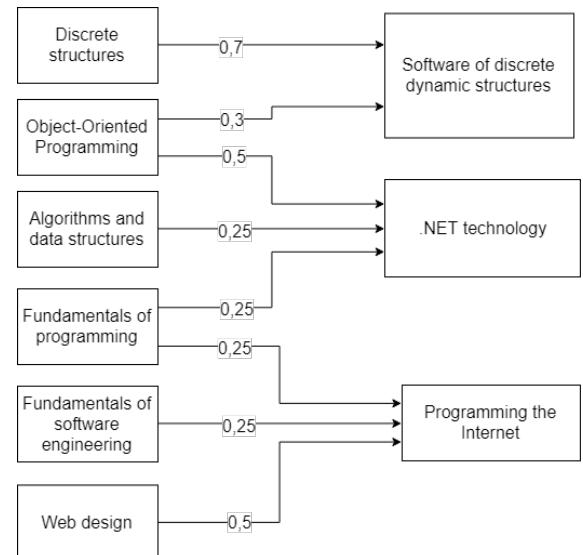


Fig. 1. A fragment of the structural-logical scheme of the related disciplines of the specialty "Software Engineering"

After forming a set of all possible combinations of the formed groups (subgroups), we calculate the value for them  $A_{j,g,m+1}$  (on initial iterations  $\varphi_{j,m}$  and then determine the best combination by the equation (1). The time spent on forming all possible combinations of the subgroup depends on the number of groups (subgroups) and the number of students in these groups (subgroups). Reduction the timing of these operations may be the subject of a study in the future.

Note, during the program implementation of equation (1), it is advisable to select several "best" combinations of groups (subgroups) and give the user the opportunity to select the "optimal" combination. Formation of subgroups of students in all subsequent semesters will be carried out on the basis of the results of previously studied adjacent disciplines, asking them with the coefficients of importance, based on the structural-logical scheme of the specialty. For example, in Figure 1, a fragment of the structural-logical scheme of adjacent disciplines for the specialty "Software Engineering".

The equation for determining the knowledge potential of a subgroup based on the results of related disciplines corresponds to equation (4). The equation (3) will be used to determine the knowledge

For example, the formation of subgroups for the discipline ".NET Technology" will be based on the results of previous related disciplines, where, according to the equation (3)

$z_1 = 0,5$  - the coefficient of the importance of discipline "Object-oriented programming",  $z_2 = 0,25$  - "Algorithms and data structures",  $z_3 = 0,25$  - "Fundamentals of programming". Student's points for the above-mentioned disciplines will be, for example,  $p_1 = 75$ ,  $p_2 = 77$ ,  $p_3 = 86$  accordingly. Then, according to equation (3), his knowledge potential in relation to the discipline ".NET Technology" will equal:

$$\varphi_{j,k,0} = 0,5 * 75 + 0,25 * 77 + 0,25 * 86 = 78,5$$

Now we can formulate an algorithm for the implementation of the proposed method for forming the composition of academic groups (subgroups) with the use of a diffusion-like model, whose block diagram is shown in Figure 2.

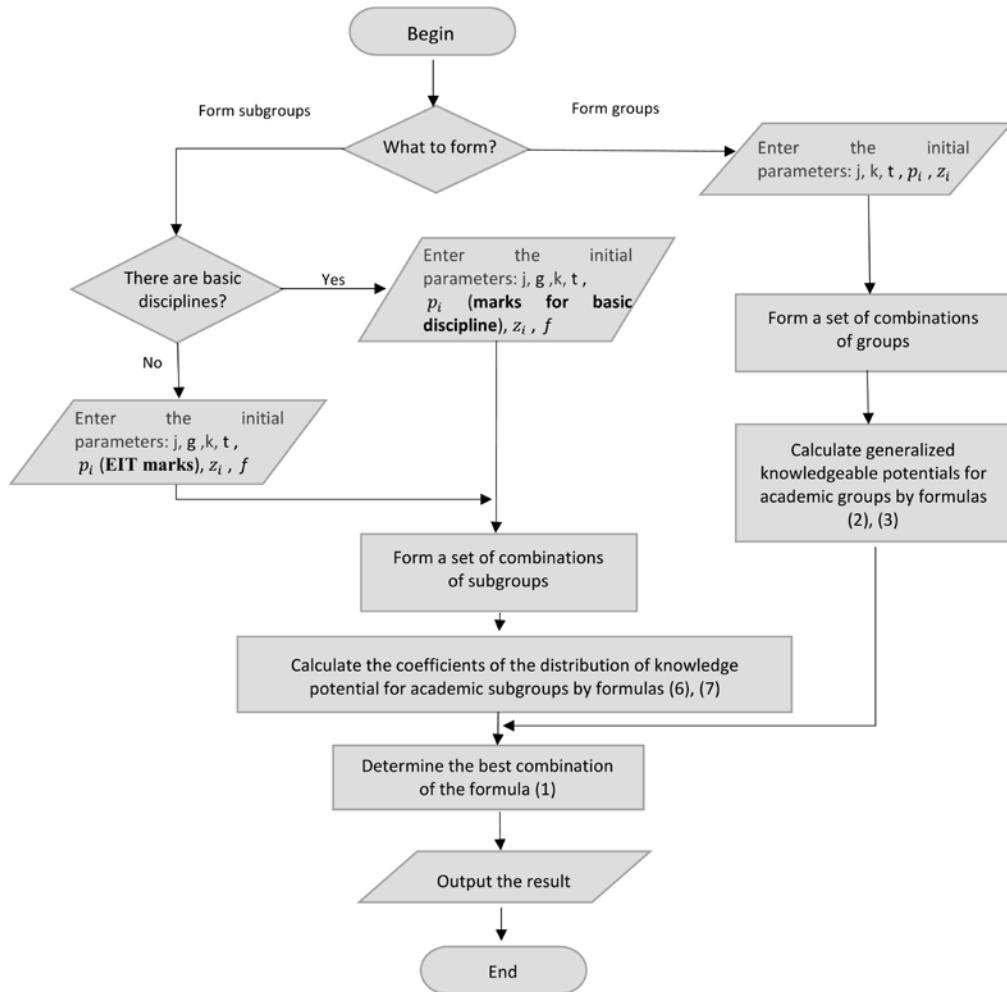


Fig.2. Algorithm for the implementation method for forming the composition of academic groups (subgroups) with the use of a diffusion-like model

### III. SOFTWARE REALIZATION

The software system of forming the composition in academic groups based on a diffusion model is developed using an object-oriented approach and .NET technology, programming language C#. Student data, lists of formed

groups, as well as other data necessary for the functioning of the system are stored in RDBMS MySQL.

Consider functional details of system. After launching the program, the user will be given full access to the entire functional system. The "Form groups" function gives the user the ability to form groups of first-year students based on the

results of external testing. In this case, each item will be given a coefficient of importance that corresponds to the chosen specialty.

The logical and conceptual description of the functionality system for the function "Form groups" is reflected in the sketch of the form, which is presented in Figure 2.

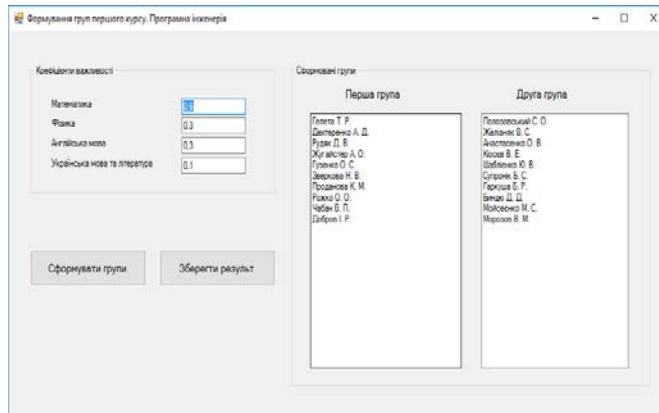


Fig. 3. Program's window for forming a composition of groups

The function "Form subgroups" is similar to the "Form groups" function, but in this case, possible combinations will be formed as a result of division of the group into subgroups. The division into subgroups will be carried out for a specific discipline, taking into account the results of previously passed disciplines, which are the basis for this discipline in accordance with the structural-logical scheme of the specialty.

In general, to split a group into a subgroup, you must first select from the list the course, then the group that is on this course. Then choose the discipline for which the subgroup will be subdivided. After selecting the discipline there will be a list of disciplines that are the basis or adjacent to the discipline for which division is carried out.

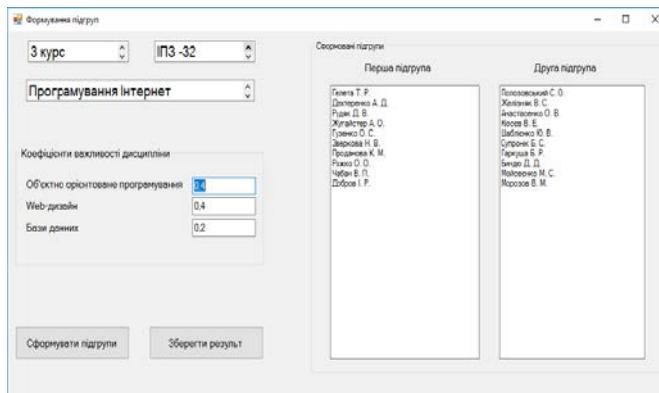


Fig. 4. Program's window for forming a composition of subgroups

The lecturer of discipline serves as a source of knowledge for the formed subgroup and affects its know-how. Therefore, for him, the coefficients of efficiency will be given, which will reflect his level of knowledge, which we will determine based on the acquired qualification level (table 1).

The logical and conceptual description of the functionality of the system for the function "Form subgroups", is reflected in the sketch of the form, which is presented in Figure 3.

Once the groups and sub groups have been formed, the user will be able to save them or form them again, having previously changed the parameters.

#### IV. CONCLUSION

The paper analyzes the existing methods of forming the composition of academic groups (subgroups) and shows that they are not effective, because they do not take into account the influence of the interaction of students between groups and the lecturer on the effectiveness of the educational process. It is shown that for constructing the method of forming the composition of academic groups it is expedient to use the diffusion model of the process of dissemination of knowledge potential.

A new method is proposed for the formation of composition of academic groups based on a diffusion-like model using the coefficient of distribution of the knowledge potential of the group as the main comparison parameter.

#### REFERENCES

- [1] I. M. Avdeeva, "Innovative communicative technologies in the work of the curator of the academic group". Kyiv, Ukraine: *Professional*, 2007, 304 p. (in Ukrainian)
- [2] E. Bern, "Leader and group. On the structure and dynamics of organizations and groups". Yekaterinburg, Russia: *Litur*, 2000, 320 p. (in Russian)
- [3] L. A. Dzyuba, "Psychological factors the introduction of modern educational technologies in higher education," in *Trends and modern psychological and pedagogical problems of preparation of specialists in high school*. Lugansk, Ukraine: Publishing House of Volodymyr Dahl East-Ukrainian National University, 2002, pp. 101-102. (in Ukrainian)
- [4] R. L. Krichevsky, "Psychology of a small group: theoretical and applied aspects". Moscow, Russia: *MSU*, 1991, 205 p. (in Russian)
- [5] G. Middleman, "Sozialarbeit group," in *Encyclopedia of social work*, vol. 3. Moscow, Russia: Center for Universal Human Values, 1994. 465 p. (in Russian)
- [6] "External independent evaluation," Oct. 2008. [Online]. Available: <http://www.osvita.ua>. [Accessed: Dec.3,2017]. (in Ukrainian)
- [7] A. Y. Bomba, M. V. Nazaruk, V. V. Pasechnik "Building diffusion model the information propagation process by a knowledgeable potential," in *Bulletin of National University "Lviv Polytechnic"*, no. 800, Computer science and information technology, 2014, pp. 35-45.(in Ukrainian)
- [8] M. V. Nazaruk and V. V. Pasechnik, "Simulation of the urban educational environment as a relevant social network" in *Information technologies and computer engineering*, vol. 3, Vinnitsa, Ukraine: VNTU, 2013, pp. 42-47. (in Ukrainian)

# The Mobile Environment Monitoring System with a Web Interface

Nataliia Furmanova, Galina Shilo, Anton Kalynychenko, Pavlo Kostianoi

Department "Information technology of electronic devices", Zaporizhzhia National Technical University, UKRAINE, 69063,  
Zaporizhzhia, Zhukovskogo str., 64, [shilo.gn@gmail.com](mailto:shilo.gn@gmail.com), [nfurmanova@gmail.com](mailto:nfurmanova@gmail.com)

**Abstract:** In the scientific work the mobile system has been developed that will allow to remotely control the level of atmospheric air pollution. The statistic data of air pollution are shown at the developed web site. These statistics is allowed to forecast the air pollution at the district

**Keywords:** air pollution, environment monitoring, ecological device, software complex, mobile system.

## I. INTRODUCTION

Environmental monitoring is complex observations of the environment, including components of the natural environment, natural ecological systems, processes in them, phenomena, assessment and changes forecast of the environment.

The development of monitoring as a complex method of the gathering information about the observed object and its activity analysis allows to talk about the formation of an information monitoring technology that combines diagnostics (assessment of the current), genesis (assessment of the past), and the forecast (assessment of the future) of the state of the studied objects [1].

This article proposes the development of a software and hardware complex that will allow remote control over the level of air pollution. This complex can be used by installing remote monitoring modules at multiple points to control emissions produced by industry and cars and quickly provide control environmental conditions for a rapid response structures of civil defense and labor protection in enterprises and government agencies.

## II. FORMULATION OF THE PROBLEM

Issues of environmental monitoring are relevant today. This is evidenced by a large number of scientific papers on this topic. Thus, in [2, 3] environmental monitoring is considered as a part of Smart Cities and the importance of control with air content in urban cities is emphasized.

Different mathematical models are used for solving the task of modeling air pollution. In [4] forecast methods of the ecological situation on the basis of a hidden Markov model are specified. Using the interval difference operators for analysis of air pollution from vehicular traffic is proposed in [5]. Wireless sensor networks using Wi-Fi are proposed in [6], but this decision has some constraints, so we are proposed using SIM modem.

## III. THE MAIN FEATURES

Following steps realize the processing of information in the developed monitoring system:

- measurement of gases in the air;
- processing information by microcontroller;
- recording results on the server;
- presenting results on the site;
- view the given information by end-users.

Existing air pollution monitoring devices do not have the ability to transmit data for their analysis remotely.

Based on the considered analogs, we have put forward and implemented the following requirements:

### A. Modularity

To be able to use complex on different objects, cities, countries-complex must have universal measurements. This flexibility has been achieved by using a series of sensors MQ, which allow you to measure a huge range of gases without changing the polling algorithm.

### B. Measurement ranges

It can be achieved the following ranges of measurements of different gases with a combination of different sensors:

- LPG and propane: 200ppm-5000ppm
- Butane: 300ppm-5000ppm
- Methane: 5000ppm-20000ppm
- H2: 100ppm-10000ppm
- Alcohol: 100ppm-2000ppm
- CH4 and natural gas: 200-10000ppm
- LNG and iso-butane: 200-1000ppm
- Carbon-monoxide: 20ppm-2000ppm
- Ozone: 10-1000ppm
- Ammonia, Benze, Hydrogen: 10ppm-10000ppm
- H2S: 1ppm-200ppm
- Ammonia: 5ppm-500ppm
- Toluene, Acetone, Ethanol: 5ppm-500ppm

### C. Data gathering

The complex provides remote collection of information from a variety of monitoring modules located at a great distance from each other. The collection is carried out by transferring information from each module to the server using the SIM800L GSM modem. This solution has some advantages in comparison with the using of Wi-Fi. Firstly, mobile communication covers significantly larger territories. Secondly, in the absence of communication device will write data in the memory and send them later.

### D. Presentation of information

The ability to display information in the form of interactive graphs using any device with Internet access is implemented, as well as daily, weekly, monthly sampling data with information about exceedances.

Reporting is carried out by a web application, where you can select the data for the certain period of time (Fig. 1) and look through them in the form of interactive graphs, as well as a table (Fig. 2).

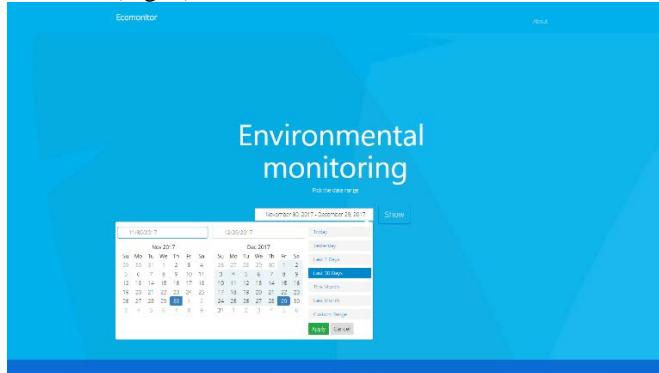


Fig. 1. The page of selecting data samples for the period of time.

#### E. Energy consumption

It implements the ability to work from the power supply 5 V, as well as due to the very low energy consumption from the battery.

#### F. Price

Since reviewed analogues are rather expensive, they cannot be used in combination. Therefore, one of the main tasks was to maximally reduce the cost of individual monitoring modules.

This problem was solved due to relatively cheap components: a microcontroller (PIC16f887), gas sensors of the MQ series and a temperature and humidity sensor DHT.

Due to this we were able to reduce the cost of the device to approximately 750 UAH.

### III. HARDWARE

#### A. The microcontroller

There are following requirements to the microcontroller: at least 3 analog-to-digital converters, low power consumption, built-in UART interface, low cost. Based on this, the PIC16F887 microcontroller was chosen.

#### B. The gas sensors

As sensors for determining the gas concentration optimally take sensors MQ-X series by FC-22. The main advantage of this series is that identical in function to the sensors ensure the measurement of the concentration of a whole range of gases. It allows to measure several parameters, as well as to interchange the sensors without changing the polling algorithm.

Table I lists all kinds of compatible sensors.

#### C. GSM modem

As a remote data transmission module, the GSM/GPRS modem SIM800L is used. The standard SIM800L control interface provides access to GSM/GPRS 850/900/1800 / 1900 MHz network services for sending calls, SMS messages and exchanging digital GPRS data. The module is controlled via the UART interface using AT commands.

The SIM800L component has an implemented TCP / IP protocol stack, automatic detection of the AT command

control rate, sending and receiving GPRS data (TCP / IP, HTTP, etc.).

#### D. Temperature and humidity sensor

As a temperature and humidity sensor it was optimal to choose one product in order to reduce the cost and dimensions of the device. As a combined humidity and temperature sensor, we chose the DHT22 digital sensor. The main characteristics of the sensor are ultra-low power consumption, lack of tying, long life time, digital interface.

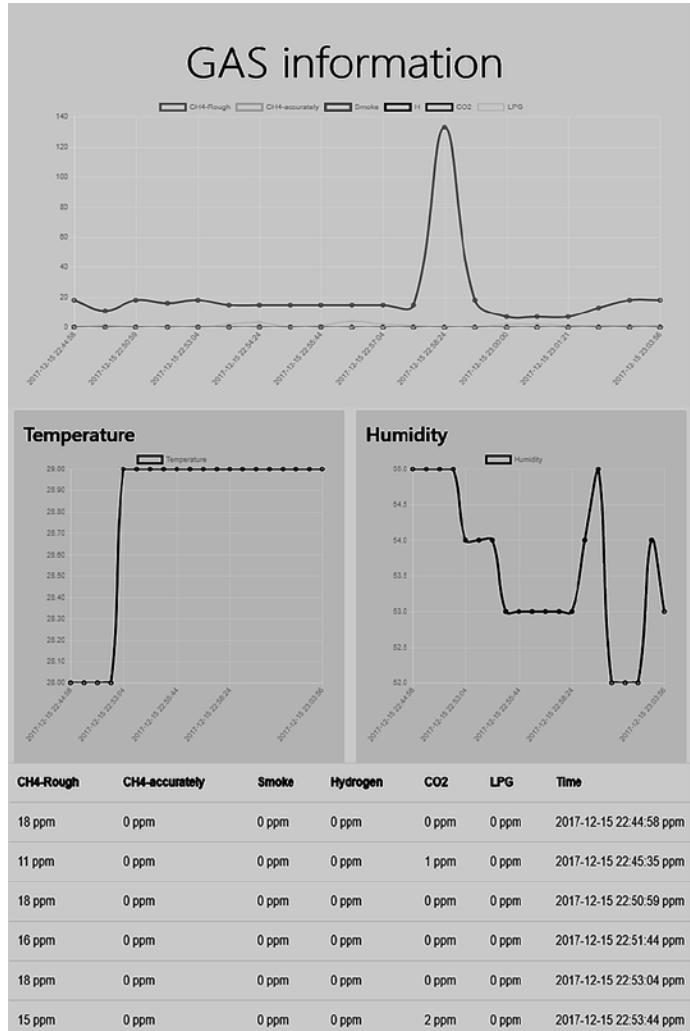


Fig. 2. The page with information of air pollution.

#### E. Printed circuit board (PCB)

Due to the use of surface mounting elements and plating of the board, we were able to place all components on a single-sided PCB with dimensions of 50x22 mm.

Tracing of the printing unit was carried out on the basis of an electric schematic diagram in the easyEDA system. As a result, a template was created and a photorealistic image of the printed circuit board was obtained (Fig. 3).

As a material of the designed PCB heat-resistant glass fiber was chosen. Its thickness is 1.5 mm, it has copper oxide foil, 50 mkm.

The shape of the printed circuit board is rectangular; the board is fastened using a threaded connection. Radioelements are fixed on the board by soldering.

TABLE 1. MQ SERIES

| Model | Target Gas  |
|-------|---|
| MQ-2  | General combustible gas   |
| MQ-3B | Alcohol   |
| MQ-4  | Natural gas, Methane  |
| MQ-5B | LPG, Natural gas, Coal gas                                      |
| MQ-6  | LPG, Propane  |
| MQ-7B | Carbon Monoxide (CO)  |
| MQ-8  | Hydrogen  |
| MQ-9B | CO and Combustible gas  |
| MQ131 | Ozone O <sub>3</sub>  |
| MQ135 | Air Quality Control (NH <sub>3</sub> , Benzene, Alcohol, smoke) |
| MQ136 | Sulfureted Hydrogen (H <sub>2</sub> S)                          |
| MQ137 | Ammonia (NH <sub>3</sub> )                                      |
| MQ138 | VOC (Mellow, Benzene, Aldehyde, Ketone, Ester)                  |

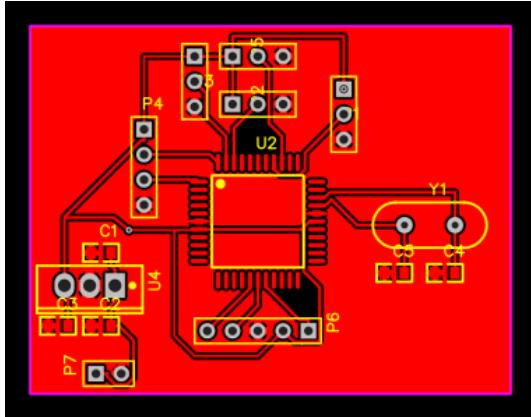


Fig. 3. Photorealistic image of the PCB.

The appearance of the printed circuit board is shown at the Fig. 4.

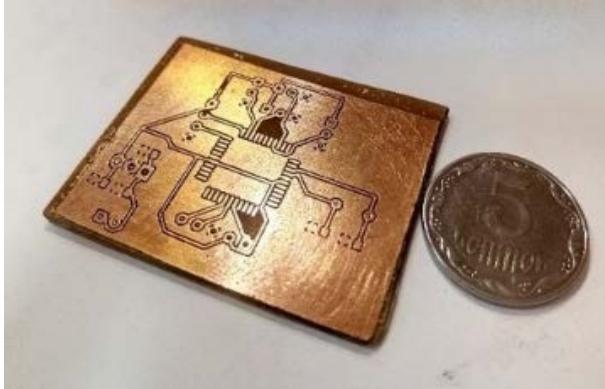


Fig.4. The appearance of the produced PCB.

#### F. The housing of device

For the developed PCB of the device, we designed a rectangular shaped housing made of plastic. Since the model

is developed as a prototype, to test the housing and the device as a whole, housing parts were manufactured by 3D printing using a 3D printer for FDM technology. PLA (polylactide) based on environmental materials was chosen as the material. Diameter of the used plastic fibers is 1.75 mm, which is due to the features of the printer structure. A three-dimensional model was created in the SolidWorks system (Fig. 5).

The housing consists of four parts: the base, partitions with sensor mount and a lid. The overall dimensions of the assembled housing are 70x70x40 mm. After testing it is possible to produce a series of pressure casting housings, that will reduce the cost of the housing unit as a whole in mass production.

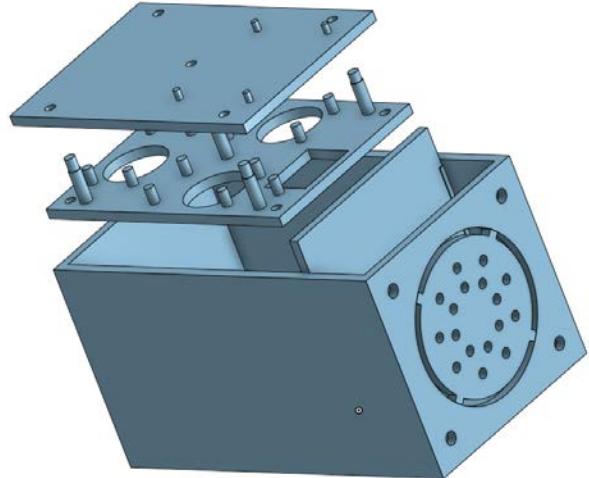


Fig. 5. The housing model without a lid.

## IV. SOFTWARE

### A. The software part of a module

When the module is turned on, the microcontroller calibrates the MQ-X gas sensors, and then it initializes the connection to the GSM module. After that, the cyclic starts measure the level of air pollution, the information is gathered from the sensors and sent it to the server. The time between measurement and sending can be changed from 10 seconds to several days.

During the delay between the measurements, the controller switches to sleep mode, which reduces power consumption.

If the connection to the GSM module has not been established, the microcontroller will write the data into memory and retry the sending of data at the next cycle pass.

Sending data to the server occurs through the HTTP POST request. This allows to record the readings in the form of a JSON string and to process on the server side easy.

### B. The Server

The work of the complex is to monitor a large area and collect information from a variety of modules. Therefore, it is necessary to store, process and submit a fairly large flow of information, so we chose the following software tools:

- 1) MySQL - Database management system, which is used to store information received from modules.
- 2) WEB server (in particular nginx HTTP server + php)
  - using REST requests, we can use it both to receive

information from the modules and display information to the end user.

### C. Server Software

When the SIM800 modem installed in the module is accessed to the server, the script parses the incoming JSON string to an array of data from each sensor, as well as information about the module, and places these data to the database.

Information about the database tables used for the operation of the complex is given below (Table II, Table III).

TABLE 2. TABLE DEVICES

| Column name | Information                              |
|-------------|--|
| Id          | Record number (table key)                |
| Address     | The physical address of the installation |
| Name        | The device name                          |
| Description | Device description / Notes               |

TABLE 3. TABLE READINGS – INFORMATION FROM SENSORS

| Parameter | Meaning                        |
|-----------|--------------------------------|
| Id        | Record number (table key)      |
| Dev_id    | ID of device                   |
| Datetime  | Time and date of data fixation |
| Temp      | Temperature from DHT22 sensor  |
| Humidity  | Humidity DHT22 sensor          |
| Module1   | Sensor 1 readings              |
| ...       | ...                            |
| Module9   | Sensor 1 readings              |

### D. User Interface

When designing the complex, the ability to view information from any device in an easy and understandable way has provided. Therefore, for these purpose it was decided to create a web interface and present data in the form of interactive graphs. Experimental results are shown in [7].

The web application was written in javascript, which makes a selection for a given period of time and displays this data using the library chart.js.

## V. CONCLUSION

In this paper, a system for environmental monitoring is proposed. A module has been developed that will allow to monitor the level of air pollution remotely.

The low cost of the developed module will make it possible to produce and install a large number of modules for environmental monitoring to cover extended areas.

The following tasks were solved. The automation system for determining the concentration of gases in the atmosphere by means of sensors was studied. An algorithm for the

functioning of the device is developed. A layout and trace of the printed circuit board were created; performance characteristics were calculated. The technology of production of parts and assemblies was chosen an experimental model was created. A system for storing, processing and reporting information on air pollution in a user-friendly form was developed.

Analysis of the collected information on the state of the environment, taking into account rose winds allows to determine the source of emissions of certain gases; to simulate environmental conditions and predict the movement of air currents in the event of emergencies.

Features of the developed ecological monitoring system:

- a wide range of air pollution coverage;
- the possibility of modification or individual assembly on request of the customer;
- small, in comparison with similar devices, cost;
- relatively simple process of production using typical technological processes;
- gathering information from a set of modules located at different points.

These properties allow using a proposed system for educational purposes. It was reanalyzed at the Zaporizhzhia National Technical University in form of laboratory stand. This stand was developed to study the basics of data analysis by students. The stand allows them to vary the components of the modules, test their operation in practice and improve its features using modern approaches for gathering and analysis of data.

## REFERENCES

- [1] J. Awange, and J.B. Kyalo Kiema, Environmental Geoinformatics: Monitoring and Management, New York, Springer, 2013, 533 p.
- [2] H. Song, R. Srinivasan, T. Sookoor and S. Jeschke, Smart Cities: Foundations, Principles, and Applications, Wiley; Hoboken, NJ, USA: 2017.
- [3] M. Bacco; F. Delmastro; E. Ferro and A. Gotta "Environmental Monitoring for Smart Cities," *IEEE Sensors Journal*, Vol. 17, Issue: 23, 2017, pp. 7767 – 7774
- [4] Wei Sun, Hao Zhang, Ahmet Palazoglu et al., "Prediction of 24-hour-average PM2.5 concentrations using a hidden Markov model with different emission distributions in Northern California", *Science of the Total Environment*, pp. 93-103, 2013.
- [5] I. Voytyuk, N. Porplitsya, A. Pukas and T. Dyvak "Identification the interval difference operators based on artificial bee colony algorithm in task of modeling the air pollution from vehicular traffic," *14th International Conference The Experience of Designing and Application of CAD Systems in Microelectronics (CADSM)*, 2017, pp. 58-62.
- [6] M. Pavani, P.T. Rao, "Real time pollution monitoring using Wireless Sensor Networks", *2016 IEEE 7th Annual Information Technology Electronics and Mobile Communication Conference (IEMCON)*, pp. 1-6, 2016.
- [7] Statistics of air pollution monitoring: <http://ecomon.devus.tk>

# Visitors Queue Management Optimization using Web System for Activity Support of the Administrative Services Center

Oleksandr Papa, Yevhen Kedrin, Andriy Pukas, Ruslan Avhustyn

Department of Computer Science, Ternopil National Economic University, UKRAINE, Ternopil, 8 A. Chekhova str, email:  
[poa@tneu.edu.ua](mailto:poa@tneu.edu.ua), [place.ev@gmail.com](mailto:place.ev@gmail.com), [apu@tneu.edu.ua](mailto:apu@tneu.edu.ua), [r.avhustyn@tneu.edu.ua](mailto:r.avhustyn@tneu.edu.ua)

**Abstract:** In this paper the improving of quality and minimizing the waiting time at managing the queue of visitors during providing services in the Administrative Services Center of Ternopil city is described.

**Keywords:** City Council, Administrative Services Center, Queue Optimization, Web System.

## I. INTRODUCTION

The Ternopil City Council (TCC) provides a large number of services in various spheres of its activities (land relations, housing, passport, communal, etc.). TCC department, which coordinates these services, defined as Administrative Services Center (ASC) [1]. Taking into account the number of Ternopil population and a significant number of legal entities, a large number of people visit the ASC every day. It leads to queue forming that require be managing and optimizing.

The tasks of queue managing and optimization is well investigated for structured queues [2, 3].

The queue in ASC now is formed using an electronic display and terminal, which allows viewing information about available services and getting a ticket with the number in queue. That is kiosk-based queue. Advising and inquiries processing are handled by the administrators which are qualified employees assigned to a specific group of services. Servicing in the center is in three halls:

Hall 1 - Sector for providing administrative services;

Hall 2 - Sector for providing permitting procedures;

Hall 3 - Sector for providing nonadministrative services.

For registering in queue, it is necessary to know in what hall the selected service is providing. Further client need to register himself and receive a ticket for servicing.

In existing system when registering in a queue, using the terminal, a new number is immediately assigned to one of the administrators.

The administrator is chosen randomly, so several employees can get a different number of clients. This approach does not take into account the duration of servicing, so customers have to wait for assigned administrator, even when the other is free and could provide the service. Therefore, the situations are possible when the customer is registered the first one, but waits longer.

On the other hand, in ASC the web-based support system is functioning which allows online registration for ASC servicing [4, 5]. Whereby, it became necessary to combine and provide service for two queues of customers – people, who were registered online and those ones, who were registered in terminals in ASC. Nowadays, the division of the

queue is made when an one administrator services only online registered clients. However, in this case, an administrator should wait for a client, who did not come in time but he could provide services for those who registered in queue using terminals.

There are registration and information terminals providing the list of services in ASC. However, these terminals do not show full important information, which is on ASC web site. To make changes or add some information require separate changing for a site and for a terminal that is inconvenient and ineffective for administrators.

Therefore, web-based support system for ASC queue management should be upgraded.

## II. MODELING SYSTEM REQUIREMENTS

Nowadays, existing ASC web-based support system allows online registration. However, the existing functional should be extended by module that combining and processing online queue and terminal-based queue. The queue list should be formed for a present day for an administrator and the list should be shown only for his hall. Registration numbers or time in case of online registration should be shown in this list. For each position in this list it should be shown the current status, for example "Waiting for" or "Is now servicing by an administrator №". For starting and stopping servicing, an administrator should have the possibility to press the keys "Start servicing" and "Stop servicing". If a customer does not come in time, there should be functions to move records down or to take off from queue.

Software for terminals and display panels should be updated. It allows showing selected web pages accessible only in ASC network. Queue list will be shown on display panel for each hall. Number of ticket, which are servicing just now will be displayed by green color, and which waiting for services by yellow color, and all other by grey color. Creeping line with useful information will be placed at the bottom of screen.

Information terminal will contain information about services, working hours in the center, contacts and registration on the next days etc. A registration terminal will be used for a registration where it is necessary to select a hall and receive a ticket with an indicated number in the queue. At the same time, for online registration it is necessary to choose a service or hall, day and time. For a registration it is necessary to indicate personal surname, name and patronymic, phone number, number of tasks to solve and confirm personal agreement for processing of personal data.

Having successfully registered, a customer has a possibility to print a ticket for receiving a service. A registered date and time are considered as reserved and blocked for other customers.

At software modeling considering above mentioned requirements and using MVC approach, the database has been developed. Its general structure and relations are presented in class diagram (Fig. 1).

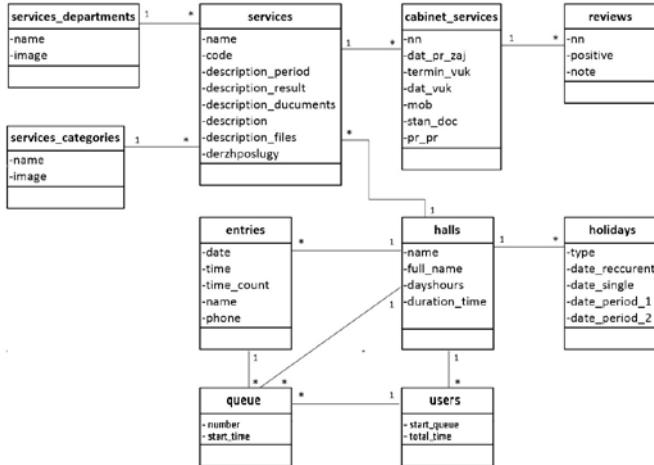


Fig.1. Class diagram of proposed system

The proposed structure provide the possibility to take into consideration functional and non-functional system requirements.

### III. SYSTEM IMPLEMENTATION

Taking into account that the existing web-based support system was developed in PHP software language, MySQL was used as DBMS, dynamic and standardized interfaces were built using jQuery and Bootstrap and page output on devices with different resolution is made using CSS Media Queries, it is the same tools are proposed to use for modules compatibility and support of the corresponding processing speed [3].

The set of technical tools, which is used for formation and management of the queue in ASC, was presented in Fig. 2.



Fig. 2. Set of used technical tools in ASC

Therefore, the possibility to view and manage the queue was added into an administrative interface to meet specified requirements. Interfaces for terminals and display panels have been developed on the base of web site in combination with special software for pages viewing and surfing.

A page for the queue processing was added into an administrative interface. In Fig. 3, the list of customers, who registered for the hall 1 for current day at the given moment, was shown. All records from the last day are automatically

archived. An administrator can see only his customers. The number and time of registration, and the status of processing were indicated in the list.

| Visitor     | Information | Status                        |
|-------------|-------------|-------------------------------|
| Ticket # 8  | from 11:30  | Served by the administrator 6 |
| Ticket # 9  | from 11:31  | Served by the administrator 8 |
| Ticket # 10 | from 11:32  | Waiting                       |

Fig. 3. Queue display in an administrator's profile

To start servicing, an administrator should visit the administrative profile of the site and press the button "Start servicing". After this, an administrator will see the number of a customer, who should come to him (Fig. 4). After providing the service, an administrator should press the button "The service provided, close the ticket", after this the next customer from the queue will be directed to an administrator.

| Visitor     | Information | Status                        | Action                                 |
|-------------|-------------|-------------------------------|--|
| Ticket # 10 | from 11:32  | Served by the administrator 6 | The service provided, close the ticket |
| Ticket # 8  | from 11:30  | Served by the administrator 8 |  |
| Ticket # 9  | from 11:31  | Served by the administrator 8 |  |
| Ticket # 13 | from 11:33  | Waiting                       |  |

Fig. 4. Processing of a customers queue

If an administrator needs to have a break, he can press the button "Stop servicing". Then customers will be directed to another administrator and if the present customer has not been serviced, he will return to the queue and will be directed to the next free administrator. There are some additional menu selection which are opened while pressing the button "In addition" (Fig. 5).

| Information | Action                          |
|-------------|---------------------------------|
| from 11:32  | Close ticket and stop servicing |
| from 11:30  | Move to 1 position              |
| from 11:31  | Move to 2 positions             |
| from 11:33  | Move to 3 positions             |
|             | Take off from the queue         |

Fig. 5. Menu items of the queue

In case, an administrator serviced a customer's request and wants to stop servicing, the menu item "Close the ticket and

stop servicing” can be selected. As well, an administrator can use the button “The service provided, close the ticket” and after this “Stop servicing”. However, in this case, the next customer will be directed to an administrator.

In case, if a customer did not come in time he can be moved in some positions down in the queue or if he did not come he can be taken off from the queue. In Figure 6 a customer with the number 10 was moved in one position and the next customer was directed to an administrator. The status of the moved customer is changed into “Waiting, changed position”.

| Visitor     | Information | Status                        |
|-------------|-------------|-------------------------------|
| Ticket # 13 | from 11:33  | Served by the administrator 2 |
| Ticket # 8  | from 11:30  | Served by the administrator 6 |
| Ticket # 9  | from 11:31  | Served by the administrator 8 |
| Ticket # 10 | from 11:32  | Waiting, changed position     |
| Ticket # 14 | from 11:46  | Waiting                       |

Fig. 6. Moving a customer in the queue

In addition to customers, who were registered using the terminals, also in the queue list are included customers, who online registered using ASC web site (Fig. 7). Unlike terminal registrations, those, who were online registered, have not number but time of servicing and their registered surnames, names, patronymic and phone number. They do not differ and the processing of such customers is made the same way. The list of online registration can be seen at the tab “Registration for servicing”. Such records appear in the list of the queue before the time of servicing.

| Visitor               | Information  | Status                        |
|-----------------------|--|-------------------------------|
| Ticket # 13           | from 11:33   | Served by the administrator 2 |
| Ticket # 9            | from 11:31   | Served by the administrator 8 |
| Ticket # 10           | from 11:32   | Waiting, changed position     |
| Ticket # 14           | from 11:46   | Waiting                       |
| Ticket # 15           | from 11:50   | Waiting                       |
| Online queue<br>12:00 | John Doe<br>0987654321<br>Question: One<br>Created: 14-04-2018 | Waiting                       |

Fig. 7. Online registered customers on display

In Figure 8, the form of administrator editing is presented, where can be found his hall and working place for servicing. If only hall was indicated, it can be seen only the list of the queue. The path to edit the site users: Admin section – Users – Manager – Select the user – Tab “ASC”.

Fig. 8. Hall and working place of an administrator

Information and registration terminals with updated functional are located in the center. In Figure 9, the homepage of the information terminal, which is formed on the base of the web site functional, is shown. Here, users can see all necessary information and register online for the next day.



Fig. 9. Homepage of the information terminal

In Figure 10, the web page of the information terminal with working hours of the center was shown.

You are here: [Home page](#) ▶ [About the center](#) ▶ [Work schedule](#)

## Work schedule

Monday, Wednesday - from 8.00 to 17.15  
 Tuesday, Thursday - from 8.00 to 20.00  
 Friday - from 8.00 to 16.00  
 Saturday - from 8.00 to 15.00 (except for real estate)

## Schedule of clients servicing

Monday, Wednesday, Friday - from 9.00 to 16.00  
 Tuesday, Thursday - from 11.00 to 20.00  
 Saturday - from 8.00 to 15.00 (except for real estate)

Fig. 10. Working hours of the center

In Figure 11, the example of web page of the information terminal with the service list was shown.

You are here: Home page > Services > Register all services

| <a href="#">Services by categories</a>   | <a href="#">Services by departments</a> | <a href="#">All services</a> | <a href="#">Sort by halls</a> | <a href="#">Sort by name of A-Z</a> | <a href="#">Sort by name Z-A</a> |
|--|---|------------------------------|-------------------------------|-------------------------------------|----------------------------------|
| Search by name or code...  |   |                              |                               |                                     |                                  |
| Act of inspection of specialized enterprises and / or their reception points for compliance with the requirements of the Law of Ukraine "On Scrap Metal"         | Hall 1                                  |                              |                               |                                     |                                  |
| Introduction of non-residential buildings, structures, non-residential premises in residential buildings to the city's housing stock                             | Hall 3                                  |                              |                               |                                     |                                  |
| capture of citizens for apartment registration at the place of residence of the executive committee of the city council, removal from the apartment registration | Hall 3                                  |                              |                               |                                     |                                  |
| Registration of citizens wishing to enter a housing cooperative, registration  | Hall 3                                  |                              |                               |                                     |                                  |
| Production of copying on a scale of 1: 500, 1: 2000  | Hall 3                                  |                              |                               |                                     |                                  |
| Issuance (reissue, issue of duplicate and cancellation) of conclusions on the approval of land management projects for land allocation                           | Hall 2                                  |                              |                               |                                     |                                  |

Fig. 11. Screen of information terminal with the service list

In Figure 12, the homepage of the registration terminal was presented. Now, while registration there is no need to select a concrete service, it is enough to select the hall where it is provided.

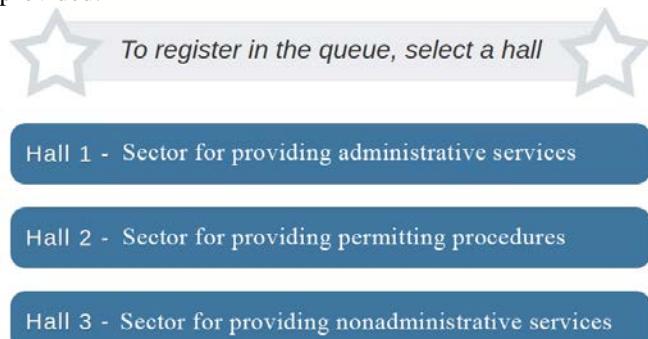


Fig. 12. Screen of registration terminal

Having selected the hall, a customer is proposed to confirm his selection pressing the button "Print a ticket" (Fig. 12). It was developed to protect wrong registration.

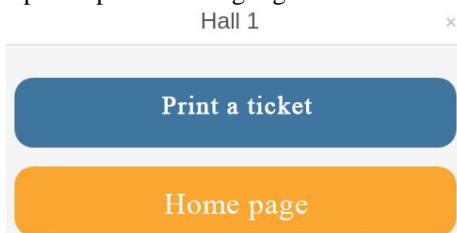


Fig. 13. Confirmation of registration

In Figure 14, the sample of the ticket printed by the registration terminal is shown.



Fig. 14. Registration ticket

In Figure 15, the interface of a display panel of the first hall was shown. Registrations, which are being processed at the moment by an administrator is displayed by green light. Those registrations, which are in the queue and will be processed in a while, are displayed by yellow light. The number of such registrations depends on the number of administrators, who are providing services for customers at the given moment.



Fig. 15. Display panel with registration list

Having finished the servicing, a number of an administrator disappears from the display panel, and the number of the next one will change its color from yellow one into green and it is blinking for some time changing the text and informing about the number of the servicing table. In case of online registrations not registration number is shown but the time of the registration and for some time the text will be changed from the time into "Online queue".

#### IV. CONCLUSIONS

Research results anticipate their use to support servicing of individuals and legal entities registered in Ternopil in ASC at TCC. As result the software was developed which supports enhance customer service and improve efficiency by optimizing the queue management.

The software for queue management in ASC has been modeled and the functional of web site, terminals and display panels are combined to provide optimization.

The interfaces of software, registration and information terminals, display panels have been developed; their role and peculiarities have been described in the article.

Proposed applied results may be used for solving similar queue optimization tasks.

#### REFERENCES

- [1]. "Regulations on the Administrative Service Center in Ternopil city", Ternopil, 2013, 3 p.
- [2]. Asmussen S. *Applied Probability and Queues*, Springer-Verlag New York, 2003, 438 p.
- [3]. Tijms H.C. *A First Course in Stochastic Models*, Wiley, Chichester, 2003, 478 p.
- [4]. Papa O.A., Kedrin Ye.S., Pukas A.V. "Features of implementation the web-based system for support the Administrative Services Center". *Materials of the All-Ukrainian Conference with International Participation "Modern Computer Information Technologies", ACIT'2017*, Ternopil, TNEU, 2017, p. 167 - 169.

# Perspective-Correct Computation Pixels Color for Systems of Three-Dimensional Rendering

Oleksandr Romaniuk<sup>1</sup>, Oleksandr Dudnyk<sup>1</sup>, Serhii Pavlov<sup>2</sup>, Olena Tsikhanovska<sup>3</sup>

1. Department of Software Engineering, Vinnytsia National Technical University, UKRAINE, Vinnytsia, 95 Khmelnytske shose str., email: [tom8591@gmail.com](mailto:tom8591@gmail.com), [dudnyk@vntu.edu.ua](mailto:dudnyk@vntu.edu.ua)
2. Department of Biomedical Engineering, Vinnytsia National Technical University, UKRAINE, Vinnytsia, 95 Khmelnytske shose str., email: [psv@vntu.edu.ua](mailto:psv@vntu.edu.ua)
3. Department of Economics of Enterprises and Corporations, The Vinnytsya Educational and Research Institute of Economics Ternopil National Economic University, UKRAINE, Vinnytsia, 37 Gonta str., email: [vnnie.epik@gmail.com](mailto:vnnie.epik@gmail.com)

**Abstract:** In computer graphics, the projections of three-dimensional images are considered on a two-dimensional picture plane. Flat geometric projections are divided into two main classes: central and parallel. The difference between them is determined by the relationship between the center of the projection and the projection plane. If the distance between them is finite, then the projection will be central, and if it is infinite, then the projection will be parallel. In real space reflection of rays from objects is perceived at the location of the observer, that is, on the principle of central projection. Correct reproduction of colors takes place provided that the components of the color intensities of the corresponding surface points in the global (object) and screen coordinate systems coincide.

The authors propose methods to improve the performance of texture mapping and realism of shading, in particular, the methods perspective-correct texturing and Phong shading.

**Keywords:** texturing, texture mapping, rendering, computer graphic, shading, Phong shading.

## I. INTRODUCTION

When forming graphical images is solved by a twofold task – improve performance and enhance realism. Today the performance of graphical tools sufficient for the formation of images according to their visual properties similar to the photos, you have achieved photographic quality.

To ensure high realism is important to consider the perspective transformation of polygons in the determination of pixel colors: in the texture mapping and shading.

The perspective-correct formation of colors is used both for shading and texture mapping.

When coloring the surfaces of three-dimensional objects, the methods of Gouraud and Phong are most often used. The question of the prospective correct reproduction of colors according to these methods is considered, respectively, in [1, 2, 3].

In the tasks of texturing, you need to find the relationship between the screen coordinates and texture coordinates. In order to ensure high productivity, the linear and quadratic functions are often used in perspective-correct texture mapping [4, 5]. In such approaches, a molded image may have artifacts and does not always faithfully reproduce the

perspective of an object. In order to increase the realism of perspective-correct texturing [4, 5, 6], use of nonlinear functions, the calculation of which involves the implementation of labor-intensive operations.

## II. SHADING WITH THE PERSPECTIVE

Ignoring the depth of the object in the calculation of the vectors leads to error computing its orthogonal components, which can be calculate by the formula

$$\begin{aligned} \Delta I &= I_A + (I_B - I_A) \cdot \frac{u \cdot z_I}{z_2 - u \cdot (z_2 - z_I)} - I_A - \\ &- (I_B - I_A) \cdot u = (I_B - I_A) \cdot \left( u - \frac{u \cdot z_I}{z_2 - u(z_2 - z_I)} \right) = \\ &= (I_B - I_A) \cdot u \cdot \left( 1 - \frac{1}{\frac{z_2}{z_I} + u(1 - \frac{z_2}{z_I})} \right), \end{aligned} \quad (1)$$

replacing the intensity value of the color value of the orthogonal component.

For perspective-correct reproduction of colors using Phong shading it is necessary to use non-linear interpolation of normal vectors using a variable  $t_w$ . Unfortunately, the calculation  $t_w$  according to the formula [1]

$$t_w = \frac{Z_{Aw} \cdot t_v}{Z_{Bw} - t_v(Z_{Bw} - Z_{Aw})} \quad (2)$$

provides for the execution of a division operation for each current value of  $t_v$ . Consider the approximation of  $t_v$  to simplify the hardware implementation. Since the dependence is nonlinear, using linear interpolation on the whole interval variable is excluded. Approximation  $t_w$  second degree polynomial  $a \cdot t_v^2 + b \cdot t_v + c$ . Find unknown  $a, b, c$ . To do this we set up a system of equations using three points  $t_v = 0, t_v = 1, t_v = 1/2$ .

$$\begin{cases} c = 0, \\ a + b + c = 1, \\ \frac{1}{4} \cdot a + \frac{1}{2} \cdot b + c = \frac{Z_{Aw}}{Z_{Bw} + Z_{Aw}}. \end{cases}$$

The system has such a solution

$$a = \frac{2 \cdot (Z_{Bw} - Z_{Aw})}{(Z_{Bw} + Z_{Aw})}, \quad b = \frac{(3 \cdot Z_{Aw} - Z_{Bw})}{(Z_{Bw} + Z_{Aw})},$$

$$c = 0.$$

$$\text{If } \tilde{\lambda} = \frac{Z_{Bw}}{Z_{Aw}}, \text{ then } a = \frac{2 \cdot (\tilde{\lambda} - 1)}{(\tilde{\lambda} + 1)}, \quad b = \frac{(3 - \tilde{\lambda})}{(\tilde{\lambda} + 1)}.$$

The quadratic approximation gives satisfactory results only for  $\tilde{\lambda} \leq 3$ . In figure 1 shows a graph of change of the absolute error of the approximation from  $t_v$ ,  $\tilde{\lambda}$ .

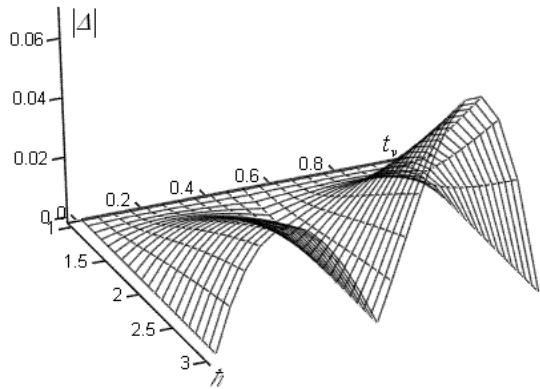


Fig. 1. The dependence of the modulus of the absolute

error of the approximation from  $t_v$ ,  $\tilde{\lambda}$

Higher accuracy of approximation can be achieved if the use of piecewise quadratic interpolation on two periods of change  $t_v$ . For  $0 \leq t_v \leq 0,5$

$$a = \frac{8 \cdot Z_{Aw} \cdot (Z_{Bw} - Z_{Aw})}{(Z_{Bw} + Z_{Aw})(3 \cdot Z_{Aw} + Z_{Bw})},$$

$$b = \frac{(3 \cdot Z_{Aw} + Z_{Bw})}{(Z_{Bw} + Z_{Aw})(3 \cdot Z_{Aw} + Z_{Bw})}, \quad c = 0.$$

For  $0,5 < t_v \leq 1$

$$a = \frac{-8 \cdot Z_{Bw} \cdot (Z_{Aw} - Z_{Bw})}{(Z_{Bw} + Z_{Aw})(3 \cdot Z_{Aw} + Z_{Bw})},$$

$$b = \frac{2 \cdot (9 \cdot Z_{Aw} - 5 \cdot Z_{Bw}) \cdot Z_{Bw}}{(Z_{Bw} + Z_{Aw})(3 \cdot Z_{Aw} + Z_{Bw})},$$

$$c = \frac{3 \cdot (Z_{Aw} - Z_{Bw})^2}{(Z_{Bw} + Z_{Aw})(3 \cdot Z_{Aw} + Z_{Bw})}.$$

The analysis showed that in this case  $\tilde{\lambda} = 2, 3, 4, 5$  the maximum modulus of the relative error does not exceed, 1%, 4%, 8%, 13%. With regard to three-dimensional objects,  $\tilde{\lambda}$ , as a rule, does not exceed 3.

Consider using approximation by third-order polynomial of the form  $a \cdot t_v^3 + b \cdot t_v^2 + ct + d$ . For finding the unknown we set up a system of four equations. To do this, make the value of the polynomial equal  $t_w$  (see formula 2) in points  $t_v = 0, 1/3, 2/3, 1$ . Find:

$$a = \frac{9 \cdot (Z_{Bw} - Z_{Aw})^2}{(2 \cdot Z_{Bw} + Z_{Aw}) \cdot (Z_{Bw} + 2 \cdot Z_{Aw})},$$

$$b = \frac{-9 \cdot (Z_{Bw} - Z_{Aw})(Z_{Bw} - 2 \cdot Z_{Aw})}{(2 \cdot Z_{Bw} + Z_{Aw}) \cdot (Z_{Bw} + 2 \cdot Z_{Aw})},$$

$$c = \frac{(2 \cdot Z_{Bw}^2 - 4 \cdot Z_{Aw} \cdot Z_{Bw} + 11 \cdot Z_{Aw})}{(2 \cdot Z_{Bw} + Z_{Aw}) \cdot (Z_{Bw} + 2 \cdot Z_{Aw})}.$$

The analysis showed that when using the cubic interpolation achieves better accuracy compared to piecewise-quadratic interpolation. For example, when  $\tilde{\lambda} = 2, 3, 4, 5$  the maximum modulus of the relative error does not exceed, 0,64 %, 2,9 %, 6,3 %, 10,6 %.

### III. IMPROVING THE PERFORMANCE OF TEXTURING WITH PERSPECTIVE

Finding texture coordinates is a time consuming procedure because it requires the execution of complex operations for each pixel according to the formula [4, 5].

$$u = \frac{Ax + By + C}{Gx + Hy + I}, \quad v = \frac{Dx + Ey + F}{Gx + Hy + I}.$$

If  $x_{i+1} = x_i + I$ , then

$$u_{i+1} = \frac{A_j \cdot (x_i + I) + B_j \cdot y_i + C_j}{D \cdot (x_i + I) + E \cdot y_i + F} = \frac{(A_j \cdot x_i + B_j \cdot y_i + C_j) + A_j}{(D \cdot x_i + E \cdot y_i + F) + D}.$$

For  $u_0$  the formula has the form:

$$u_0 = \frac{A_1 \cdot (x_0 + 0) + B_1 \cdot y_i + C_1}{D \cdot (x_0 + 0) + E \cdot y_i + F} = \frac{A_1 \cdot x_0 + B_1 \cdot y_i + C_1}{D \cdot x_0 + E \cdot y_i + F}.$$

For  $u_1$ :

$$u_1 = \frac{A_1 \cdot (x_0 + 1) + B_1 \cdot y_i + C_1}{D \cdot (x_0 + 1) + E \cdot y_i + F} = \frac{A_1 \cdot x_0 + A_1 + B_1 \cdot y_i + C_1}{D \cdot x_0 + D + E \cdot y_i + F}.$$

For  $u_2$ :

$$\begin{aligned} u_2 &= \frac{A_1 \cdot (x_0 + 2) + B_1 \cdot y_i + C_1}{D \cdot (x_0 + 2) + E \cdot y_i + F} = \\ &= \frac{A_1 \cdot x_0 + A_1 \cdot 2 + B_1 \cdot y_i + C_1}{D \cdot x_0 + D \cdot 2 + E \cdot y_i + F}. \end{aligned}$$

Thus, for  $u_n$  the formula has the form:

$$\begin{aligned} u_n &= \frac{A_1 \cdot (x_0 + n) + B_1 \cdot y_i + C_1}{D \cdot (x_0 + n) + E \cdot y_i + F} = \\ &= \frac{A_1 \cdot x_0 + A_1 \cdot n + B_1 \cdot y_i + C_1}{D \cdot x_0 + D \cdot n + E \cdot y_i + F}. \end{aligned} \quad (2)$$

A similar formula can be written for  $v_n$ .

$$v_n = \frac{A_2 \cdot x_i + B_2 \cdot y_0 + B_2 \cdot n + C_2}{D \cdot x_i + E \cdot y_0 + E \cdot n + F}.$$

From the given formulas it is visible that for calculation of each texel computer needs to perform 2 operations of division, 8 operations of addition and 8 multiplications.

As can be seen from formulas (2), the value of the expressions  $A_1 \cdot x_0 + B_1 \cdot y_i + C_1$  and  $D \cdot x_0 + E \cdot y_i + F$  for each  $n$  remain unchanged, and therefore can be calculated once for each rasterization line using formula:

$$u_t = A_1 \cdot x_0 + A_1 \cdot n + B_1 \cdot y_i + C_1, \quad u_b = D \cdot x_0 + E \cdot y_i + F.$$

where  $u_t$  and  $u_b$  constant part of the numerator and denominator of formula (1) in accordance.

Therefore,  $u_n$  can be calculated according using the formula

$$u_n = \frac{u_t + A_1 \cdot n}{u_b + D \cdot n}. \quad (3)$$

Thus, for calculating coordinates of all texels, variables,  $u_{1t}$  and  $u_{1b}$  are constant, and their value is sufficient to calculate once. Similarly you can calculate  $v_n$ .

This simplification reduces the number of operations of addition and multiplication to 4 for each.

Based on these mathematical relationships it is possible to offer algorithm of parallel calculation of texels coordinates: for each rasterization line at first calculates the parameters  $u_t$  and  $u_b$ , then calculate values of the numerator and denominator in parallel by the formula:

$$w_n = u_t + A_1 \cdot n, \quad v_n = u_b + D \cdot n,$$

then the division operation to calculate the coordinates

$$u_n = \frac{w_n}{v_n}.$$

The sequence of operations of the algorithm shown in figure 2.

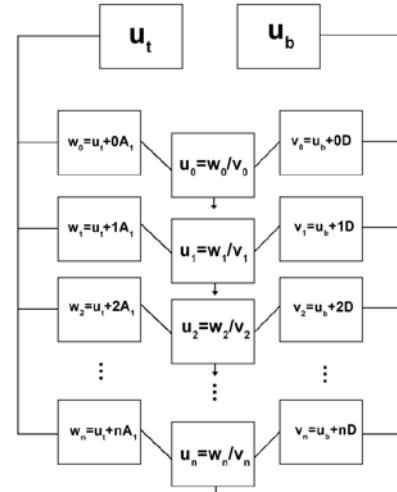


Fig. 2. Parallel calculating texel coordinates

Since for each  $x$ ,  $n$  is increased by 1, there is such a formula:

$$w_n = w_{n-1} + A_1, \quad v_n = v_{n-1} + D. \quad (4)$$

In accordance with the formulas (4), we can offer a consistent algorithm for calculating texture coordinates:  $w_n$  and  $v_n$  for each  $x$  calculated by adding to  $w_{n-1}$  and  $v_{n-1}$ , that was calculated for  $x-1$ ,  $A_1$  and  $D$ . The sequence of operations of the algorithm shown in figure 3.

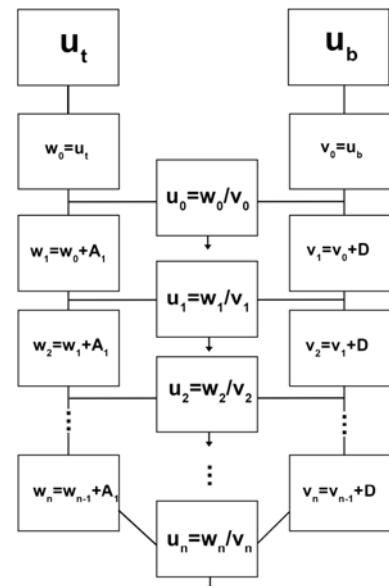


Fig. 3. Sequential calculating texel coordinates

The computing of texture coordinates by this algorithm can also be accelerated using parallel computing. One possible approach to parallelization is the parallel rasterization of several lines simultaneously. However, this approach will not be productive enough in cases when the number of pixels per row is much higher than the number of rows. Therefore, it is advisable to use means of parallel calculations within a single line.

Let's consider parallel computing of cordiant texels for pixels located at even and odd positions in the rasterization line (fig. 4).

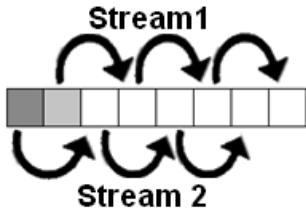


Fig. 4. Parallel computing texels coordinates for two threads on the same rasterization line

If  $w_n = w_{n-1} + A_1$ , a  $v_n = v_{n-1} + D$  then:

$$w_{n+1} = (w_{n-1} + A_1) + A_1, \quad v_{n+1} = (v_{n-1} + D) + D.$$

Thus, parallel computation of the texture coordinates of pixels on odd and even positions is possible according to the formulas:

$$w_n = w_{n-2} + 2A_1, \quad v_n = v_{n-2} + 2D. \quad (5)$$

In this case, the coordinates texels for the first two points are defined by the formula (3).

Based on the formulas (5) and (5) to establish the relationship which allows parallel rasterization the line in an random number of threads using formulas:

$$w_n = w_{n-k} + kA_1, \quad v_n = v_{n-k} + kD,$$

whre  $k$  – the number of concurrent threads.

It is also possible a parallel calculation of the coordinates in two streams by simultaneous rasterization of line in two directions (fig. 5). From right to left according to the formula (4), and from left to right according to the formula:

$$w_n = w_{n+1} - A_1, \quad v_n = v_{n+1} - D.$$

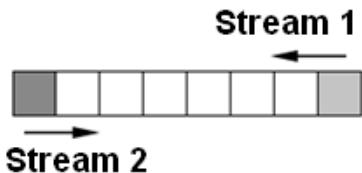


Fig. 5. Parallel computation of the coordinates in the two streams with a counter direction of rasterization

The figure 6 show that the proposed method makes it possible to increase productivity perspective-correct texturing 26%. Testing was conducted on the Intel i7 2600K CPU and GPU AMD RX460.

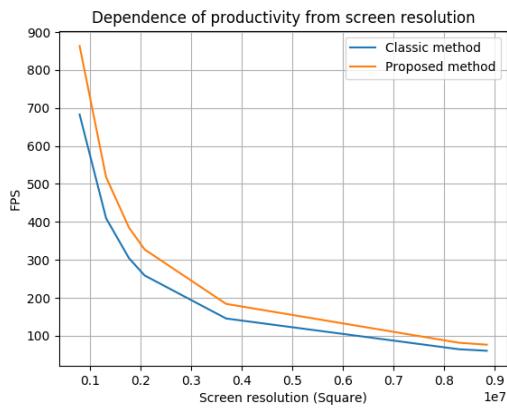
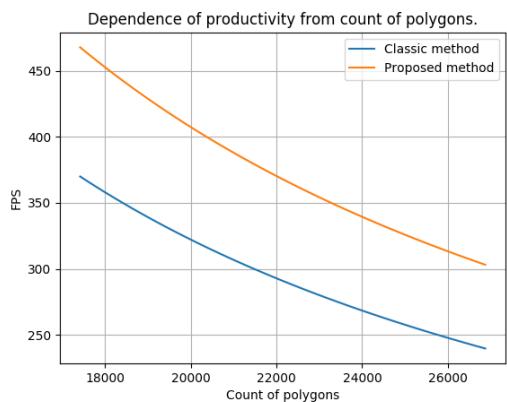


Fig. 6. Comparison of the performance of the proposed method with the classical

### III. CONCLUSION

Proposed methods to improve the performance of texture mapping and realism of shading, in particular, the methods perspective-correct texturing and Phong shading.

### REFERENCES

- [1] G. F. Ahmed, R. Barskar, J. Bharti, and N. S. Rajput, "Content Base Image Retrieval Using Fast Phong Shading," 2010 International Conference on Computational Intelligence and Communication Networks, 2010.
- [2] R. F. Lyon, "Phong Shading Reformulation for Hardware Renderer Simplification", *Apple Technical Report #43*, August 2, 1993.
- [3] D. A. Kulagin, "Models of shading. Flat model. Shading on Guro and Fong ", *Computer graphics. Theory, algorithms, examples on C++ and OpenGL*. [Online]. Available [http://compgraphics.info/3D/lighting/shading\\_model.php](http://compgraphics.info/3D/lighting/shading_model.php)
- [4] P. S. Heckbert, "Survey of Texture Mapping," *IEEE Computer Graphics and Applications*, vol. 6, no. 11, pp. 56–67, 1986.
- [5] F. Tsai and H. C. Lin, "Polygon-based texture mapping for cyber city 3D building models," *International Journal of Geographical Information Science*, vol. 21, no. 9, pp. 965–981, 2007.
- [6] X. U. Ying, "An Improved Texture Rendering Technique Based on MipMap Algorithm", *Journal of Mianyang Normal University*, 2013, 5: 017.

# GPU-Based Rendering for Ray Casting of Multiple Geometric Data

Sergei Vyatkin<sup>1</sup>, Oleksandr Romaniuk<sup>2</sup>, Oleksandr Dudnyk<sup>2</sup>

1. Synthesizing Visualization Systems Laboratory, Institute of Automation and Electrometry SB RAS, RUSSIAN FEDERATION, Novosibirsk, 1 Akademika Koptyuga avenue, email: [sivser@mail.ru](mailto:sivser@mail.ru)
2. Department of Software Engineering, Vinnytsia National Technical University, UKRAINE, Vinnytsia, 95 Khmelnytske shose str., email: [tom8591@gmail.com](mailto:tom8591@gmail.com), [dudnyk@vntu.edu.ua](mailto:dudnyk@vntu.edu.ua)

**Abstract:** We present a new GPU-based rendering method for ray casting of multiple geometric data. Our approach supports a polygonal data to calculate scattered light. Terrain is represented for the base of scalar perturbation functions. The geometric model is based on non-polygonal representation. We present a new representation scheme for freeform surfaces it is possible to combine basic surface and perturbation functions. To recognize the place it is often sufficient to fill areas with the generalized texture patterns, such as "themes". Themes designed beforehand and consumption of memory for storing them is not huge.

**Keywords:** freeform surfaces, recursive multilevel ray casting, shape texture, thematic texture, scattered light, volume-oriented visualization.

## I. INTRODUCTION

Using traditional polygonal representation for the example complex surface give rise to a range of problems such as visible surface determination, depth complexity handling, controlling levels of details, clipping polygons by viewing frustum, geometry transformations of large number of polygons [1, 2].

A method for constructing a triangle mesh whose vertices coincide with the zero-valued isosurface is the Marching Cubes algorithm [3]. Although it provides many greater capabilities, the use of voxel-based terrain in real-time virtual simulations also introduces several new difficulties. The algorithms used to extract the terrain surface from a voxel map produce far greater numbers of vertices and triangles when compared to conventional 2D terrain. The development of a seamless LOD algorithm for voxel-based terrain is vastly more complex than the analogous problem for height-based terrain.

Texturing and shading of voxel-based terrain is more difficult than it is for height-based terrain. In the cases that triangle meshes are generated for multiple resolutions, arises the cracking problem.

A method for patching cracks on the boundary plane between cells triangulated at different voxel resolutions was described in [4]. Using a voxel-based model however, can achieve the same results at a much lower hardware requirement.

The proposed method includes the following main features: rendering second-order surfaces; rendering surfaces defined on a regular altitude grid (Shape texture); rendering freeform surfaces; scattered light visualization.

## II. VOLUME-ORIENTED RENDERING

Along with possibility to rasterizing 2D space, the main feature of the proposed method is rasterizing made by 3D space quadtree subdivision of pyramids of different levels, which constitute the whole pyramid of vision. Then a pyramid of the lowest level is binary-tree subdivided into voxels of the lowest level - Recursive Multilevel Ray Casting (RMLRC). In the latter case, extent space regions (in depth) can masked out. Depth subdivision of space performed on the logarithmic basis. The technique of RMLRC allows determining an intersection of a ray (pyramid) of any level with a surface effectively. It is also suitable for fast culling of a spatial region outside an object. The core of this approach is effective search of volume elements (hereinafter voxels), involved in current frame generation, fused with direct projection [5]. If geometry transformation is described by matrix (C) then new calculated matrix of quotients  $(Q)'=(C)T^*Q^*(C)$  does in the coordinate system P the same as the matrix (Q) in the coordinate system P. The matrix (C) of projecting transformation calculated once for particular frustum. So the use of projecting transformation generalizes the discussed algorithm for pyramidal volumes (frustums) and allows synthesize images with perspective. In our approach a virtual environment can be described using polygonal models, surfaces of second order, three-dimensional scalar functions, defined on discrete grid  $h=h(u,v)$ , free-form surfaces, which are represented by composition of base surfaces and shape-driving functions. The proposed RMLRC algorithm has several advantages in processing mentioned surface models as compared with known algorithms. Therefore, for example, RMLRC algorithm applied for second order surfaces simplifies calculation of Phuong shading, since at the last level of subdivision derivative plane coefficients for the surface obtained. Further, the possibility of second order and free form surfaces composition allows producing objects that are more complex. Photorealistic visualization of complex surfaces, for instance, specified by a three-dimensional function, defined on discrete grid  $h=h(u,v)$  can be done without intermediate polygonal approximation. Such surfaces represented by differential height map, i. e. the algebraic carrier surface is given and in each grid node, only deviation from this surface needed. This representation facilitates calculation of contiguous levels of details as well as makes easy quality filtering. Geometry transformations apply only to the carrier surface and the height map is kind of surface

texture. Freeform surface representation differs from known representations, such as for example, Constructive Shell Representation [6], since the former is free of problems arising when complex surfaces are approximated with large number of Bezier patches, B-splines etc. (problems of rendering patches boundaries, problems of warped halfspaces). The proposed approach allows specifying surface representation as composition of base surfaces and shape-driving functions. Small number of such functions is enough to describe surfaces of any form including non-convex, with holes etc. (Fig. 1). Discussed surface representations and the approach to their processing facilitate description of such phenomena as waves, dynamic surface warping, morphing, deformations and animation of wide range of surfaces.



Fig. 1. Freeform surface (F-117) over terrain. Height map resolution 512x512. Screen resolution 1920 x 1080

### III. FREEFORM SURFACES

Traditionally, the parametric form represents each patch in a freeform surface as a mapping from 2D parameter space to 3D space. Although parametric patches are powerful for constructing freeform surfaces, processing these patches poses fundamental problems, only two - constructive solid geometry (CSG) and boundary representation (Brep) - commonly represent solids exactly, that is, without approximation. In particular, "separation" and other such problems associated with curved halfspaces are hard to solve in parametric patches. We present a new representation scheme for freeform surfaces - it is possible to combine basic surface  $F(x,y,z)$  and shape-driving function  $R(x,y,z)$  or perturbation function, where shape-driving function represents smooth deformation of basic surface [7]. The shape-driving function is composed of several second-order functions using logic intersection and union operations, it is recursively evaluated during subdivision itself, and therefore, computation of resulting  $R(x, y, z)$  is minimal. At the same time, it is moderate in computations, so it can be implemented in hardware to carry on realistic object outlook.

Because of application of the 3D space, subdivision algorithm the possibility appears to render surfaces, which usually consist of huge amount of patches such as freeform surfaces.

### IV. SHAPE TEXTURE

Non-regular terrain algorithms are more complex but have the potential to reduce the number of polygons that the system must process [1]. However, for many terrains with limited elevation gradient, the average expected reduction in

the image generator load is small for irregular grid as compared to regular grid. Each irregular grid node consumes much more computational time and memory than that of regular. Systems, which employ irregular grids, are very limited in the number of LODs that can be handled. In addition, it is very difficult to solve problems concerned with terrain deformation when using irregular grid (explosions, pits in the ground).

Volume oriented rendering and uniformity of object processing result in an efficient hidden surface removal and detection of spatial collisions. Chosen representation of terrain data is based on regular multi-level elevation map complemented with levels of detail [7]. This approach has several advantages, such as rapid generation and modification, efficient data storing and retrieving, over polygonal models.

### V. THEMATIC TEXTURE

Systematically analyzing geographer's work, Barr [8] has described it as operating with the "geographic matrix". For clearness, it can be thought as three-dimensional, i.e. two dimensions correspond to the geo-coordinating system, and the third one corresponds to a description list. Each element represents a "geographic fact" (Fig. 2). For a large database, it is resource-intensive to obtain and process geospecific photo-texture for the entire gaming area. It is assumed that most of the benefits of geospecific photo-texture could be attained by combining a large amount of generic photo-texture with a small amount of geospecific photo-texture.

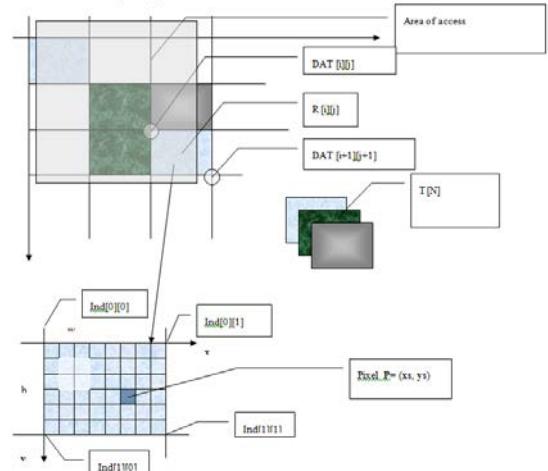


Fig. 2. Thematic texture

Simulation of big areas of terrain in the training system image generators requires the large capacity of memory for the photographic data storage. The usual approach to reduce this information is to compress the data with the help of the methods such as JPEG, etc. In the approach of the texture composition it is supposed, that the observer (pilot) often is not interested in the exact, photographic information about the areas, covered by, for example, forest, water, mountains, etc. To recognize the place it is often sufficient to fill these areas with the generalized texture patterns (we will call them "themes" afterwards). Themes designed beforehand and consumption of memory for storing them is not huge. Area of the terrain photo-texture with the same theme approximated by the polygons, which are bounded by curves. Each of the

polygons has the pointer to the texture theme and to the procedure of generating the boundaries between the areas with the different themes. The procedures of generating the boundaries do not reproduce true boundaries exactly, but have the goal to keep their features. There are two main problems in the technique of the texture composition. The first is the generation of the boundaries between the areas with the different themes. To resolve this problem, the algorithms, which allows generating the different kinds of boundaries, designed. Among the possible kinds of the boundaries the following can be selected: fractal boundaries, following natural boundaries in the maps of the texture themes (for example, following the streets in the texture map of the town), procedural blend zones (beaches, etc.). The second problem of the technique of the texture composition is the filling the areas by the homogeneous texture pattern. Filling the big area of the terrain by the small texture pattern leads to the undesirable periodic, unnatural picture. The introduction of the different kinds of non-regularity, which keeps the features of the texture pattern, allows obtaining more natural picture.

The first stage is the usual rendering of polygons, describing the texture, and filling the service buffer by the pointers to the texture themes and kinds of boundaries. The second stage is access to the buffer through the perturbation map, to obtain the color of each texel. Perturbation map calculated beforehand. Note that approach is adapted for the generation of the fractal boundaries, but it allows to use the procedural blend zones also, and to generate boundaries, which coincide with the natural boundaries in the maps. To generate fractal boundaries the following actions must be done (some are simplified): access to the perturbation map by address, defined by the U and V coordinates of the given texel, to obtain the additions to the values of U and V. Obtaining the new values of U and V by the addition with the values from the perturbation map. Access to the buffer of the source map by the address, defined by new values of U and V, to obtain the index of the texture theme; access to memory of the texture themes, to obtain the color of the texel. Boundaries between regions, obtained by the method described, looks fractal due to the fractal character of the perturbation map used. Turbulent noise used to fill this map. As it is well known, the higher spatial frequencies of the turbulent noise have the smaller amplitude (and on the contrary). All of the boundaries, obtained by this method, are similar, but different themes can have different boundaries, with different widths (amplitudes) and "crookednesses" (frequencies). Therefore, the sequence of actions, described above, must be little more complex: access to the source map to obtain the characteristics of boundary; scaling of U and V coordinates to obtain required frequency of the noise, then access to the noise map (perturbation map). Multiplication of the values, obtained from the noise map, by the constant value to provide required amplitude of the noise, then modification of U and V; access to the source map to obtain the index of the texture map; access to memory of the texture themes to obtain the color of the texel.

Obviously, blending can become correct only if textural areas have coinciding borders. To satisfy a condition an additional database is introduced which describes faces of

blending areas. These faces do not differ from others and their special role is to connect textural areas with different borders. For this purpose, such face owns border description of one area, and a texture of the other area. While the database created, these faces constructed by expanding polygons, i.e. by moving their edges. By the way, procedurally generated areas can obtained exactly the same way: if a certain polygon with any textural theme has a prescribed blended border, the procedure described above executed and a newly created polygon assigned descriptions of the area created.

The boundaries coinciding with natural borders within a textural picture are considered. Such borders can formed by processing the image from the original map buffer after all textural polygons have been rasterized. Beside difficulties with forming the borders between areas of different textures, there is a problem concerning inner filling of these areas by a homogeneous picture. Therefore, in the proposed approach the textural coordinates disturbed by a noise-map to provide singularities on the picture. It is possible to reuse noise values previously obtained while determining the textural theme index. Let us estimate memory and time consumption within the proposed approach. Note, that original map resolution can be less than final texture map resolution. Maximal spatial frequency of the final image determined by a spatial frequency of textural themes and a noise-map; therefore, they can be scaled, so that one texel of original map is transformed into an area filled by a certain picture with a fractal border. Thus, minimal size of original map determined by a required precision of border presentation. The image kept in an original map consists of contiguous areas each filled by a certain textural theme index, that is why these data can be safely compressed using, for instance, Color Cell Compression method (this will ensure up to six times compression). Besides, this task can accomplished during textural polygon rasterizing with no additional time consumption. As noise-map, contents do not explicitly emerge in the final image, periodically wrapped map can used for this purpose and unwanted regularity will not noticed. For instance, fractal border has a repeating structure only if it originally presented as a straight line parallel to one of textural coordinate axes. The task of rasterizing textured polygons and filling original map is not extremely difficult and can be handled by any sufficiently powerful universal processor, moreover the time is limited only by a needed uploading due to the camera's motion. To determine the color of a texel several steps are completed: fetch operation from original map, from noise-map, then, modification of U and V coordinates, fetch operation from original map and from textural theme memory. As one may see, these operations are quite simple and, therefore, a conventional textural fetch operation can substituted. This implementation has several significant advantages, as follows: overall memory requirements are less then those for a global texture; texture animation features become available by simple means, as fractal border can made animated, if a shift is applied to a noise-map at each frame. In addition, noise magnitude can change resulting in an interesting effect. In analogous way, texture inside the distinct regions can animated. Besides, this approach would be naturally used also for all other feature

textures (fetch operation from original texture memory should be excluded). All this could be useful while rendering, for instance, water surface, flames, snow, etc.

It should be also mentioned that implementation described above requires original map, noise-map and textural theme (as MIP map) storing. Generation of levels of detail for textural themes and noise-maps is of no difficulty, however levels of detail for original map is not clear, because these data are indices, which cannot be blended like colors. A straightforward solution of this problem is to eliminate details that are smaller than corresponding texel size, i.e. only wholly covered texels handled during rasterizing original textural polygons.

## VI. SCATTERED LIGHT

The vertex shaders compute the light reaching the eye from a source or a reflective object and fog component. The inputs to the vertex shader are vertex position, transformation matrices, sunlight intensity, the sun direction, the various extinction and scattering coefficients. In this implementation using calculate them per-vertex in a vertex shader. If the Sun is not too low in the sky, this approximation gives good results at a low cost. For these assumptions, the illumination of the ground plane and light reaching the eye from it may easily derived (Fig. 3). To find the color perceived along a line of sight, we must consider both the light reflected by objects along the line of sight, attenuated by the inverting fog, and the light scattered towards the eye by fog along the line of sight.



Fig. 3. Terrain, F117 and scattered light. Height map resolution 512x512. Screen resolution 1920 x 1080

## VII. PERFORMANCE

The visualization time is reduced by using the computational resources of a graphics processing unit with compute unified device architecture (CUDA). The result of running the programs on different processing units is the same even if they may have a different number of streaming multiprocessors. A large portion of the cube will be computed in parallel. Among the functions of the graphics processing unit was to calculate the coordinates of points of the surfaces, normals, and illumination. Geometric transformations were performed by the central processing unit (CPU). Rendering results using CPU (i7-2700K) and GPU (GTX 550Ti, GTX 750 Ti, GTX 950) are shown in Table 1.

TABLE 1. PERFORMANCE ON CPU AND GPUs

| Resolution | i7-2700K  | GTX 550Ti | GTX 750 Ti | GTX 950   |
|------------|-----------|-----------|------------|-----------|
| 256x256    | 802,65 ms | 67,03 ms  | 31,07 ms   | 26, 55 ms |
| 512x512    | 850,81 ms | 71,05 ms  | 32,93 ms   | 28, 01 ms |
| 1024x1024  | 856,52 ms | 71,53 ms  | 33,15 ms   | 30, 22 ms |

The surface obtained will be smooth (Fig. 1 and Fig. 3), and a small number of perturbation functions (16) will be necessary to create complex surface forms. The figure shows a result of modeling a scene object by means of free forms, whose description required 2Kbyte information, which is 500 times less than the polygonal description that would take 1Mbyte information.

## VIII. CONCLUSION

The main advantages include ease of calculation of points on the surface with quick search and rejection of the regions not occupied by the scene objects. A factor of 100 or more decrease in the number of surfaces for describing curved objects. Operations of the Geometry Processor (CPU) becomes significantly easier and data stream from the Geometry Processor (CPU) to the Renderer (GPU) reduced. Paralleling of the system becomes easier because the total system performance determined by the performance of the Renderer (GPU) which operations can be easily distributed between different screen regions. All problems connected to terrain generation solved. The opportunity to describe objects with grid values, i.e., with shape texture map; it becomes possible to morph objects; it is achieved by plain interpolation or animation of shape texture. It can be used to render clouds, explosions, waves on water, etc.

## REFERENCES

- [1] R. Pajarola, E. Gobbetti. "Survey on semi-regular multiresolution models for interactive terrain rendering". *The visual computer. International Journal of Computer Graphics*, Vol. 23, P. 583–605, 2007.
- [2] P. Lindstrom, D. Koller, W. Ribarsky, L. F. Hodges, F. Nick, and G. A. Turner. "Real-time, continuous level of detail rendering of height fields" *In Proceedings of Siggraph*, 1996, P. 109–118.
- [3] W. E. Lorensen, and H. E. Cline. "Marching cubes: A high resolution 3D surface construction algorithm". *Computer Graphics, Proceedings of SIGGRAPH 87*, P. 163–169.
- [4] R. Shu, C. Zhou, and M. S. Kankanhalli. "Adaptive marching cubes". *The Visual Computer*, Vol. 11, P. 202.
- [5] Vyatkin S. I., Romanyuk A. N., and Savitska L. A. "Multi-level ray casting of function-based surfaces" *Journal of Physics: Conference Series*, 803, №. 1.
- [6] P. Menon, and T. J. Watson "Constructive Shell Representation Freeform Surfaces". *IEEE Computer Graphics* 0373-17-16, 1994.
- [7] S. I. Vyatkin. "Complex Surface Modeling using Perturbation Functions". *Optoelectronics, Instrumentation and Data Processing*. 43 (3), P. 226–231 (2007).
- [8] R. Barr "Automated cartography and geographical information: Part 2", *Advances in Computer Graphics*, Volume 11, Springer, page 29, 1986.

# Estimating the Software Size of Open-Source PHP-Based Systems Using Non-Linear Regression Analysis

Sergiy Prykhodko<sup>1</sup>, Natalia Prykhodko<sup>2</sup>, Lidiia Makarova<sup>1</sup>

1. Department of Software of Automated Systems, Admiral Makarov National University of Shipbuilding, UKRAINE, Mykolaiv, Heroes of Ukraine ave. 9, email: [sergiy.prykhodko@nuos.edu.ua](mailto:sergiy.prykhodko@nuos.edu.ua)

2. Finance Department, Admiral Makarov National University of Shipbuilding, UKRAINE, Mykolaiv, Heroes of Ukraine ave. 9, email: [natalia.prykhodko@nuos.edu.ua](mailto:natalia.prykhodko@nuos.edu.ua)

**Abstract:** The equation, confidence and prediction intervals of multivariate non-linear regression for estimating the software size of open-source PHP-based systems are constructed on the basis of the Johnson multivariate normalizing transformation. Comparison of the constructed equation with the linear and non-linear regression equation based on the Johnson univariate transformation is performed.

**Keywords:** software size estimation, PHP-based system, multivariate non-linear regression analysis, normalizing transformation, non-Gaussian data.

## I. INTRODUCTION

Software size is one of the most important internal metrics of software. The information obtained from estimating the software size are useful for predicting the software development effort by such model as COCOMO II. The papers [1, 2] proposed the linear regression equations for estimating the software size of some programming languages, such as VBA, PHP, Java and C++. The proposed equations are constructed by multiple linear regression analysis on the basis of the metrics that can be measured from class diagram. However, there are four basic assumptions that justify the use of linear regression models, one of which is normality of the error distribution. But this assumption is valid only in particular cases. This leads to the need to use the non-linear regression equations including for estimating the software size of open-source PHP-based systems.

A normalizing transformation is often a good way to build the equations, confidence and prediction intervals of multiply non-linear regressions [3-5]. According [4] transformations are used for essentially four purposes, two of which are: first, to obtain approximate normality for the distribution of the error term (residuals), second, to transform the response and/or the predictor in such a way that the strength of the linear relationship between new variables (normalized variables) is better than the linear relationship between dependent and independent random variables. Well-known techniques for building the equations, confidence and prediction intervals of multivariate non-linear regressions are based on the univariate normalizing transformations, which do not take into account the correlation between random variables in the case of normalization of multivariate non-Gaussian data. This leads to the need to use the multivariate normalizing transformations.

In this paper, we build the equation, confidence and prediction intervals of multivariate non-linear regression for estimating the software size of open-source PHP-based

systems on the basis of the Johnson multivariate normalizing transformation (the Johnson normalizing translation) with the help of appropriate techniques proposed in [5].

## II. THE TECHNIQUES

The techniques to build the equations, confidence and prediction intervals of non-linear regressions are based on the multiple non-linear regression analysis using the multivariate normalizing transformations. A multivariate normalizing transformation of non-Gaussian random vector  $\mathbf{P} = \{Y, X_1, X_2, \dots, X_k\}^T$  to Gaussian random vector  $\mathbf{T} = \{Z_Y, Z_1, Z_2, \dots, Z_k\}^T$  is given by

$$\mathbf{T} = \psi(\mathbf{P}) \quad (1)$$

and the inverse transformation for (1)

$$\mathbf{P} = \psi^{-1}(\mathbf{T}). \quad (2)$$

The linear regression equation for normalized data according to (1) will have the form [4]

$$\hat{Z}_Y = \bar{Z}_Y + (\mathbf{Z}_X^+) \hat{\mathbf{b}}, \quad (3)$$

where  $\hat{Z}_Y$  is prediction linear regression equation result for values of components of vector  $\mathbf{z}_X = \{Z_1, Z_2, \dots, Z_k\}$ ;  $\mathbf{Z}_X^+$  is the matrix of centered regressors that contains the values  $Z_{1_i} - \bar{Z}_1, Z_{2_i} - \bar{Z}_2, \dots, Z_{k_i} - \bar{Z}_k$ ;  $\hat{\mathbf{b}}$  is estimator for vector of linear regression equation parameters,  $\mathbf{b} = \{b_1, b_2, \dots, b_k\}^T$ .

The non-linear regression equation will have the form

$$\hat{Y} = \psi^{-1}[\bar{Z}_Y + (\mathbf{Z}_X^+) \hat{\mathbf{b}}], \quad (4)$$

where  $\hat{Y}$  is prediction non-linear regression equation result.

The technique to build a non-linear regression equation is based on transformations (1) and (2), Eq. (3) and a confidence interval of linear regression for normalized data

$$\hat{Z}_Y \pm t_{\alpha/2,v} S_{Z_Y} \left\{ \frac{1}{N} + (\mathbf{Z}_X^+)^T \left[ (\mathbf{Z}_X^+)^T \mathbf{Z}_X^+ \right]^{-1} (\mathbf{Z}_X^+) \right\}^{1/2}, \quad (5)$$

where  $t_{\alpha/2,v}$  is a quantile of student's  $t$ -distribution with  $v$  degrees of freedom and  $\alpha/2$  significance level;  $(\mathbf{Z}_X^+)^T$  is one

of the rows of  $\mathbf{Z}_X^+$ ;  $S_{Z_Y}^2 = \frac{1}{v} \sum_{i=1}^N (Z_{Y_i} - \hat{Z}_{Y_i})^2$ ,  $v = N - k - 1$ ;

$(\mathbf{Z}_X^+)^T \mathbf{Z}_X^+$  is the  $k \times k$  matrix

$$(\mathbf{Z}_X^+)^T \mathbf{Z}_X^+ = \begin{pmatrix} S_{Z_1 Z_1} & S_{Z_1 Z_2} & \dots & S_{Z_1 Z_k} \\ S_{Z_2 Z_1} & S_{Z_2 Z_2} & \dots & S_{Z_2 Z_k} \\ \dots & \dots & \dots & \dots \\ S_{Z_k Z_1} & S_{Z_k Z_2} & \dots & S_{Z_k Z_k} \end{pmatrix},$$

where  $S_{Z_q Z_r} = \sum_{i=1}^N [Z_{q_i} - \bar{Z}_q][Z_{r_i} - \bar{Z}_r]$ ,  $q, r = 1, 2, \dots, k$ .

The confidence interval for non-linear regression is built on the basis of the interval (5) and inverse transformation (2)

$$\psi_1^{-1} \left( \hat{Z}_Y \pm t_{\alpha/2, v} S_{Z_Y} \left\{ \frac{1}{N} + (\mathbf{z}_X^+)^T \left[ (\mathbf{Z}_X^+)^T \mathbf{Z}_X^+ \right]^{-1} (\mathbf{z}_X^+) \right\}^{1/2} \right). \quad (6)$$

The technique to build a prediction interval is based on multivariate transformation (1), the inverse transformation (2), linear regression equation for normalized data (3) and a prediction interval for normalized data

$$\hat{Z}_Y \pm t_{\alpha/2, v} S_{Z_Y} \left\{ 1 + \frac{1}{N} + (\mathbf{z}_X^+)^T \left[ (\mathbf{Z}_X^+)^T \mathbf{Z}_X^+ \right]^{-1} (\mathbf{z}_X^+) \right\}^{1/2}. \quad (7)$$

The prediction interval for non-linear regression is built on the basis of the interval (7) and inverse transformation (2)

$$\psi_1^{-1} \left( \hat{Z}_Y \pm t_{\alpha/2, v} S_{Z_Y} \left\{ 1 + \frac{1}{N} + (\mathbf{z}_X^+)^T \left[ (\mathbf{Z}_X^+)^T \mathbf{Z}_X^+ \right]^{-1} (\mathbf{z}_X^+) \right\}^{1/2} \right). \quad (8)$$

### III. THE JOHNSON NORMALIZING TRANSLATION

For normalizing the multivariate non-Gaussian data, we use the Johnson translation system. The Johnson normalizing translation is given by

$$\mathbf{Z} = \boldsymbol{\gamma} + \boldsymbol{\eta} \mathbf{h} [\boldsymbol{\lambda}^{-1} (\mathbf{X} - \boldsymbol{\varphi})] \sim N_m (\mathbf{0}_m, \boldsymbol{\Sigma}), \quad (9)$$

where  $\boldsymbol{\Sigma}$  is the covariance matrix;  $m = k + 1$ ;  $\boldsymbol{\gamma}$ ,  $\boldsymbol{\eta}$ ,  $\boldsymbol{\varphi}$  and  $\boldsymbol{\lambda}$  are parameters of translation (9);  $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \dots, \gamma_m)^T$ ;  $\boldsymbol{\eta} = \text{diag}(\eta_1, \eta_2, \dots, \eta_m)$ ;  $\boldsymbol{\lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_m)$ ;  $\boldsymbol{\varphi} = (\varphi_1, \varphi_2, \dots, \varphi_m)^T$ ;  $\mathbf{h}[(y_1, \dots, y_m)] = \{h_1(y_1), \dots, h_m(y_m)\}^T$ ;  $h_i(\cdot)$  is one of the translation functions

$$h = \begin{cases} \ln(y), & \text{for } S_L \text{ (log normal) family;} \\ \ln[y/(1-y)], & \text{for } S_B \text{ (bounded) family;} \\ \text{Arsh}(y), & \text{for } S_U \text{ (unbounded) family;} \\ y & \text{for } S_N \text{ (normal) family.} \end{cases} \quad (10)$$

Here  $y = (x - \varphi)/\lambda$ ;  $\text{Arsh}(y) = \ln(y + \sqrt{y^2 + 1})$ .

### IV. THE EQUATION, CONFIDENCE AND PREDICTION INTERVALS OF NON-LINEAR REGRESSION TO ESTIMATE THE SOFTWARE SIZE

The equation, confidence and prediction intervals of non-linear regression to estimate the software size of open-source PHP-based systems are constructed on the basis of the

Johnson multivariate normalizing transformation for the four-dimensional non-Gaussian data set: actual software size in the thousand lines of code (KLOC)  $Y$ , the average number of attributes per class  $X_3$ , the total number of classes  $X_1$  and the total number of relationships  $X_2$  in conceptual data model from 32 information systems developed using the PHP programming language with HTML and SQL. Table I contains the data from [1] on four metrics of software for 32 open-source PHP-based systems.

TABLE I. THE DATA ON SOFTWARE METRICS

| $i$ | $Y$    | $X_1$ | $X_2$ | $X_3$  |
|-----|--------|-------|-------|--------|
| 1   | 3.038  | 5     | 2     | 10.6   |
| 2   | 22.599 | 17    | 7     | 7      |
| 3   | 32.243 | 21    | 13    | 4.524  |
| 4   | 16.164 | 13    | 11    | 7.077  |
| 5   | 83.862 | 35    | 24    | 6.571  |
| 6   | 24.22  | 13    | 9     | 8.077  |
| 7   | 63.929 | 35    | 19    | 8.029  |
| 8   | 2.543  | 5     | 3     | 9.4    |
| 9   | 6.697  | 5     | 5     | 7      |
| 10  | 55.537 | 25    | 14    | 8.64   |
| 11  | 55.752 | 39    | 10    | 9.077  |
| 12  | 62.602 | 30    | 17    | 7      |
| 13  | 67.111 | 23    | 22    | 14.957 |
| 14  | 2.552  | 3     | 1     | 8.333  |
| 15  | 12.17  | 10    | 5     | 3.7    |
| 16  | 12.757 | 13    | 9     | 5      |
| 17  | 5.695  | 7     | 3     | 8.429  |
| 18  | 7.744  | 9     | 6     | 9.222  |
| 19  | 7.514  | 4     | 1     | 8      |
| 20  | 11.054 | 9     | 9     | 3.667  |
| 21  | 29.77  | 17    | 15    | 3.412  |
| 22  | 11.653 | 9     | 8     | 8.778  |
| 23  | 6.847  | 5     | 4     | 3.6    |
| 24  | 13.389 | 7     | 5     | 11.714 |
| 25  | 14.45  | 12    | 6     | 16.583 |
| 26  | 4.414  | 6     | 3     | 3.667  |
| 27  | 2.102  | 3     | 1     | 3.333  |
| 28  | 42.819 | 20    | 18    | 3.5    |
| 29  | 4.077  | 4     | 2     | 9      |
| 30  | 57.408 | 33    | 14    | 9.242  |
| 31  | 7.428  | 7     | 3     | 7      |
| 32  | 8.947  | 15    | 5     | 4      |

For detecting the outliers in the data from Table 1 we use the technique based on multivariate normalizing transformations and the squared Mahalanobis distance [6]. There are no outliers in the data from Table I for 0.005 significance level and the Johnson multivariate transformation (9) for  $S_B$  family. The same result was obtained in [6] for the transformation (9) for  $S_U$  family. In [1] it was also assumed that the data contains no outliers.

Parameters of the multivariate transformation (9) for  $S_B$  family were estimated by the maximum likelihood method. Estimators for parameters of the transformation (9) are:  $\hat{\gamma}_Y = 9.63091$ ,  $\hat{\gamma}_1 = 15.5355$ ,  $\hat{\gamma}_2 = 25.4294$ ,  $\hat{\gamma}_3 = 0.72801$ ,  $\hat{\eta}_Y = 1.05243$ ,  $\hat{\eta}_1 = 1.58306$ ,  $\hat{\eta}_2 = 2.54714$ ,  $\hat{\eta}_3 = 0.54312$ ,

$\hat{\phi}_Y = -1.4568$ ,  $\hat{\phi}_1 = -1.8884$ ,  $\hat{\phi}_2 = -6.9746$ ,  $\hat{\phi}_3 = 3.2925$ ,  $\hat{\lambda}_Y = 153102.605$ ,  $\hat{\lambda}_1 = 243051.0$ ,  $\hat{\lambda}_2 = 311229.5$  and  $\hat{\lambda}_3 = 13.900$ . The sample covariance matrix  $S_N$  of the  $\mathbf{T}$  is used as the approximate moment-matching estimator of  $\Sigma$

$$S_N = \begin{pmatrix} 1.0000 & 0.9514 & 0.9333 & 0.1574 \\ 0.9514 & 1.0000 & 0.9006 & 0.1345 \\ 0.9333 & 0.9006 & 1.0000 & 0.0554 \\ 0.1574 & 0.1345 & 0.0554 & 1.0000 \end{pmatrix}.$$

After normalizing the non-Gaussian data by the multivariate transformation (9) for  $S_B$  family the linear regression equation (3) is built for normalized data

$$\hat{Z}_Y = \hat{b}_0 + \hat{b}_1 Z_1 + \hat{b}_2 Z_2 + \hat{b}_3 Z_3. \quad (11)$$

Estimators for parameters of the Eq. (11) are such:  $\hat{b}_0 = 1.02 \cdot 10^{-5}$ ,  $\hat{b}_1 = 0.56085$ ,  $\hat{b}_2 = 0.42491$ ,  $\hat{b}_3 = 0.05846$ .

After that the non-linear regression equation (4) is built

$$\hat{Y} = \hat{\phi}_Y + \hat{\lambda}_Y \left[ 1 + e^{-(\hat{Z}_Y - \hat{\gamma}_Y)/\hat{\eta}_Y} \right]^{-1}, \quad (12)$$

where  $\hat{Z}_Y$  is prediction result by the Eq. (11),

$$Z_j = \gamma_j + \eta_j \ln \frac{X_j - \varphi_j}{\varphi_j + \lambda_j - X_j}, \quad \varphi_j < X_j < \varphi_j + \lambda_j, \quad j = 1, 2, 3.$$

The prediction results by Eq. (12) for values of components of vector  $\mathbf{X} = \{X_1, X_2, X_3\}$  from Table I are shown in the Table II for two cases: univariate and multivariate normalizing transformations.

For univariate normalizing transformations (10) of  $S_B$  family the estimators for parameters are such:  $\hat{\gamma}_Y = 0.77502$ ,  $\hat{\gamma}_1 = 0.59473$ ,  $\hat{\gamma}_2 = 0.57140$ ,  $\hat{\gamma}_3 = 0.68734$ ,  $\hat{\eta}_Y = 0.44395$ ,  $\hat{\eta}_1 = 0.48171$ ,  $\hat{\eta}_2 = 0.49553$ ,  $\hat{\eta}_3 = 0.51970$ ,  $\hat{\phi}_Y = 2.063$ ,  $\hat{\phi}_1 = 2.900$ ,  $\hat{\phi}_2 = 0.900$ ,  $\hat{\phi}_3 = 3.304$ ,  $\hat{\lambda}_Y = 83.059$ ,  $\hat{\lambda}_1 = 36.695$ ,  $\hat{\lambda}_2 = 23.525$  and  $\hat{\lambda}_3 = 13.660$ . In the case of univariate normalizing transformations the estimators for parameters of the Eq. (11) are such:  $\hat{b}_0 = 3.11 \cdot 10^{-7}$ ,  $\hat{b}_1 = 0.43519$ ,  $\hat{b}_2 = 0.52239$  and  $\hat{b}_3 = 0.08546$ .

Table II also contains the prediction results by linear regression equation from [1] for values of components of vector  $\mathbf{X} = \{X_1, X_2, X_3\}$  from Table I. Note the prediction results by linear regression equation from [1] are negative for the three rows of data: 14, 19 and 27. All prediction results by non-linear regression equation (12) are positive.

Magnitude of relative error (MRE), mean magnitude of relative error (MMRE) and percentage of prediction (PRED(0.25)) are accepted as standard evaluations of prediction results by regression equations. The values of MRE for linear regression equation from [1], non-linear regression equation (12) for two cases (univariate and multivariate normalizing transformations) are shown in the Table II. The acceptable values of MMRE and PRED(0.25) are not more than 0.25 and not less than 0.75 respectively. The values of MMRE in the Table III indicate that only the value for Eq. (12) on the basis of multivariate normalizing

transformation is less than 0.25. Although all values of PRED(0.25) in the Table III are less than 0.75 nevertheless the values are greater for Eq. (12). All values of multiple coefficient of determination  $R^2$  in the Table III are greater than 0.75 but the value of  $R^2$  is greater for Eq. (12) on the basis of multivariate transformation.

TABLE II. PREDICTION RESULTS AND MRE OF REGRESSION EQUATIONS

| i  | Linear regression equation |        | Non-linear regression equation |        |                             |
|----|----------------------------|--------|--------------------------------|--------|-----------------------------|
|    |                            |        | univariate transformation      |        | multivariate transformation |
|    | $\hat{Y}$                  | $MRE$  | $\hat{Y}$                      | $MRE$  | $\hat{Y}$                   |
| 1  | 3.237                      | 0.0656 | 4.675                          | 0.5388 | 4.550                       |
| 2  | 24.142                     | 0.0683 | 19.965                         | 0.1166 | 19.990                      |
| 3  | 37.524                     | 0.1638 | 32.098                         | 0.0045 | 33.535                      |
| 4  | 25.916                     | 0.6033 | 23.171                         | 0.4335 | 21.292                      |
| 5  | 74.624                     | 0.1102 | 80.265                         | 0.0429 | 83.618                      |
| 6  | 23.224                     | 0.0411 | 20.524                         | 0.1526 | 18.901                      |
| 7  | 67.215                     | 0.0514 | 65.913                         | 0.0310 | 70.647                      |
| 8  | 4.127                      | 0.6228 | 5.789                          | 1.2764 | 5.169                       |
| 9  | 5.906                      | 0.1181 | 7.353                          | 0.0980 | 6.356                       |
| 10 | 46.843                     | 0.1565 | 42.098                         | 0.2420 | 43.126                      |
| 11 | 57.814                     | 0.0370 | 67.070                         | 0.2030 | 49.823                      |
| 12 | 56.995                     | 0.0896 | 53.497                         | 0.1454 | 56.651                      |
| 13 | 61.856                     | 0.0783 | 65.500                         | 0.0240 | 60.617                      |
| 14 | -2.395                     | 1.9384 | 2.202                          | 0.1370 | 2.447                       |
| 15 | 9.959                      | 0.1816 | 9.693                          | 0.2035 | 10.029                      |
| 16 | 21.218                     | 0.6632 | 18.682                         | 0.4644 | 18.105                      |
| 17 | 5.976                      | 0.0493 | 7.083                          | 0.2438 | 6.687                       |
| 18 | 13.991                     | 0.8067 | 12.911                         | 0.6673 | 11.301                      |
| 19 | -1.371                     | 1.1825 | 2.496                          | 0.6678 | 3.096                       |
| 20 | 15.385                     | 0.3918 | 13.301                         | 0.2032 | 12.850                      |
| 21 | 35.179                     | 0.1817 | 27.321                         | 0.0823 | 29.061                      |
| 22 | 17.045                     | 0.4627 | 15.461                         | 0.3268 | 13.268                      |
| 23 | 2.017                      | 0.7054 | 5.435                          | 0.2062 | 5.112                       |
| 24 | 11.462                     | 0.1440 | 10.367                         | 0.2257 | 8.661                       |
| 25 | 22.513                     | 0.5580 | 20.191                         | 0.3973 | 15.888                      |
| 26 | 1.630                      | 0.6307 | 5.318                          | 0.2048 | 5.260                       |
| 27 | -5.655                     | 3.6902 | 2.142                          | 0.0192 | 1.873                       |
| 28 | 43.975                     | 0.0270 | 37.967                         | 0.1133 | 38.631                      |
| 29 | 0.953                      | 0.7662 | 3.892                          | 0.0454 | 3.732                       |
| 30 | 57.164                     | 0.0043 | 53.121                         | 0.0747 | 54.381                      |
| 31 | 5.044                      | 0.3209 | 6.861                          | 0.0764 | 6.571                       |
| 32 | 16.360                     | 0.8285 | 12.934                         | 0.4456 | 14.258                      |
|    |                            |        |                                |        | 0.5936                      |

The confidence and prediction intervals of non-linear regression are defined by (6) and (8) respectively for the data from Table I.

TABLE III. VALUES OF  $R^2$ , MMRE AND PRED(0.25)

| Coefficients | Linear regression equation | Non-linear regression equation |                             |
|--------------|----------------------------|--------------------------------|-----------------------------|
|              |                            | univariate transformation      | multivariate transformation |
| $R^2$        | 0.9491                     | 0.9591                         | 0.9692                      |
| MMRE         | 0.4919                     | 0.2535                         | 0.2199                      |
| PRED(0.25)   | 0.5313                     | 0.7188                         | 0.7188                      |

Table IV contains the lower (LB) and upper (UB) bounds of the prediction intervals of linear and non-linear regressions

on the basis of univariate and multivariate transformations respectively for 0.05 significance level.

Note the lower bounds of the prediction interval of linear regression from [1] are negative for the thirteen rows of data: 1, 8, 9, 14, 15, 17, 19, 23, 24, 26, 27, 29 and 31. All the lower bounds of the prediction interval of non-linear regressions are positive. The widths of the prediction interval of non-linear regression on the basis of the Johnson multivariate transformation are less than for linear regression from [1] for the twenty rows of data: 1, 6, 8, 9, 14-20, 22-27, 29, 31 and 32. Also the widths of the prediction interval of non-linear regression on the basis of the Johnson multivariate transformation are less than following the Johnson univariate transformation for the twenty-three rows of data: 1-4, 6, 8-10, 15-18, 20-26, 28, 29, 31 and 32. Approximately the same results are obtained for the confidence interval of non-linear regression.

TABLE IV. BOUNDS OF THE PREDICTION INTERVALS

| i  | Bounds for linear regression |        | Bounds for non-linear regression |        |                             |         |
|----|------------------------------|--------|----------------------------------|--------|-----------------------------|---------|
|    |                              |        | univariate transformation        |        | multivariate transformation |         |
|    | LB                           | UB     | LB                               | UB     | LB                          | UB      |
| 1  | -8.886                       | 15.361 | 2.507                            | 15.664 | 2.053                       | 8.822   |
| 2  | 12.260                       | 36.024 | 5.800                            | 53.204 | 11.088                      | 35.207  |
| 3  | 25.530                       | 49.517 | 9.341                            | 65.987 | 19.149                      | 57.962  |
| 4  | 14.031                       | 37.802 | 6.642                            | 57.342 | 11.955                      | 37.129  |
| 5  | 61.845                       | 87.403 | 59.920                           | 84.392 | 47.603                      | 146.045 |
| 6  | 11.451                       | 34.998 | 5.956                            | 53.906 | 10.617                      | 32.866  |
| 7  | 54.797                       | 79.633 | 31.210                           | 81.247 | 40.528                      | 122.355 |
| 8  | -7.849                       | 16.103 | 2.713                            | 20.215 | 2.431                       | 9.838   |
| 9  | -5.998                       | 17.810 | 2.996                            | 26.099 | 3.097                       | 11.949  |
| 10 | 34.901                       | 58.785 | 13.397                           | 72.304 | 24.761                      | 74.346  |
| 11 | 43.606                       | 72.022 | 26.251                           | 82.571 | 26.759                      | 91.726  |
| 12 | 44.844                       | 69.146 | 19.861                           | 77.358 | 32.563                      | 97.782  |
| 13 | 47.957                       | 75.755 | 28.542                           | 81.562 | 33.153                      | 109.857 |
| 14 | -14.415                      | 9.625  | 2.084                            | 2.994  | 0.811                       | 5.262   |
| 15 | -2.080                       | 21.999 | 3.441                            | 33.425 | 5.255                       | 18.197  |
| 16 | 9.355                        | 33.081 | 5.492                            | 51.258 | 10.150                      | 31.513  |
| 17 | -5.925                       | 17.877 | 2.964                            | 24.822 | 3.336                       | 12.381  |
| 18 | 2.136                        | 25.846 | 4.145                            | 40.894 | 6.095                       | 20.095  |
| 19 | -13.374                      | 10.632 | 2.127                            | 4.916  | 1.198                       | 6.351   |
| 20 | 3.243                        | 27.527 | 4.154                            | 42.480 | 6.867                       | 23.133  |
| 21 | 22.801                       | 47.556 | 7.324                            | 63.400 | 15.978                      | 51.960  |
| 22 | 5.148                        | 28.943 | 4.693                            | 46.152 | 7.200                       | 23.590  |
| 23 | -10.093                      | 14.128 | 2.635                            | 19.103 | 2.367                       | 9.829   |
| 24 | -0.715                       | 23.638 | 3.576                            | 35.238 | 4.477                       | 15.796  |
| 25 | 9.337                        | 35.689 | 5.323                            | 56.560 | 8.396                       | 29.076  |
| 26 | -10.481                      | 13.741 | 2.621                            | 18.450 | 2.464                       | 10.048  |
| 27 | -17.916                      | 6.606  | 2.073                            | 2.648  | 0.410                       | 4.484   |
| 28 | 31.335                       | 56.615 | 10.895                           | 70.978 | 21.432                      | 68.748  |
| 29 | -11.043                      | 12.949 | 2.371                            | 12.014 | 1.575                       | 7.423   |
| 30 | 44.632                       | 69.696 | 19.170                           | 77.441 | 30.902                      | 94.883  |
| 31 | -6.838                       | 16.926 | 2.926                            | 23.959 | 3.273                       | 12.168  |
| 32 | 4.173                        | 28.547 | 4.090                            | 41.560 | 7.530                       | 26.021  |

Following [7] multivariate kurtosis  $\beta_2$  is estimated for the data on metrics of software from Table I and the normalized data on the basis of the Johnson univariate and multivariate transformations for  $S_B$  family. It is known that  $\beta_2 = m(m+2)$  holds under multivariate normality. The given

equality is a necessary condition for multivariate normality. In our case  $\beta_2 = 24$ . The estimators of multivariate kurtosis equal 28.66, 37.29 and 23.08 for the data from Table I, the normalized data on the basis of the Johnson univariate and multivariate transformations respectively. The values of these estimators indicate that the necessary condition for multivariate normality is practically performed for the normalized data on the basis of the Johnson multivariate transformation only and does not hold for other data.

## V. CONCLUSION

The non-linear regression equation to estimate the software size of open-source PHP-based systems is improved on the basis of the Johnson multivariate transformation for  $S_B$  family. This equation, in comparison with other regression equations (both linear and nonlinear), has a larger multiple coefficient of determination and a smaller value of MMRE.

When building the equations, confidence and prediction intervals of non-linear regressions for multivariate non-Gaussian data, one should use multivariate transformations.

Usually poor normalization of multivariate non-Gaussian data or application of univariate transformations instead of multivariate ones to normalize such data may lead to increase of width of the confidence and prediction intervals of regressions, both linear and nonlinear.

## REFERENCES

- [1] Hee Beng Kuan Tan, Yuan Zhao, and Hongyu Zhang, “Estimating LOC for information systems from their conceptual data models”, in *Proceedings of the 28th International Conference on Software Engineering (ICSE '06)*, May 20-28, 2006, Shanghai, China, pp. 321-330.
- [2] Matinee Kiewkanya, and Suttipong Surak, “Constructing C++ software size estimation model from class diagram”, in *13th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, July 13-15, 2016, Khon Kaen, Thailand, pp. 1-6.
- [3] D. M. Bates, and D. G. Watts, Nonlinear regression analysis and its applications. Wiley, 1988.
- [4] T. P. Ryan, Modern regression methods. Wiley, 1997.
- [5] S. B. Prykhodko, “Developing the software defect prediction models using regression analysis based on normalizing transformations”, in *Abstracts of the Research and Practice Seminar on Modern Problems in Testing of the Applied Software (PTTAS-2016)*, May 25-26, 2016, Poltava, Ukraine, pp. 6-7.
- [6] S. Prykhodko, N. Prykhodko, L. Makarova, and A. Pukhalevych, “Application of the squared Mahalanobis distance for detecting outliers in multivariate non-Gaussian data”, in *Proceedings of 14th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET)*, Lviv-Slavsk, Ukraine, February 20-24, 2018, pp. 962-965.
- [7] K. V. Mardia, “Measures of multivariate skewness and kurtosis with applications”, *Biometrika*, 57, 1970, pp. 519–530.

# Research of the Agree of Experts' Evaluations in the Estimation of Software Systems

Svitlana Krepoch<sup>1</sup>, Iryna Spivak<sup>2</sup>, Roman Krepoch<sup>3</sup>

Department of Computer Science, Ternopil National Economic University, UKRAINE, Ternopil, 8 Chekhova str., email:  
[s.krepoch@tneu.edu.ua](mailto:s.krepoch@tneu.edu.ua)<sup>1</sup>, [spivak.iruna@gmail.com](mailto:spivak.iruna@gmail.com)<sup>2</sup>, [jagmstar@gmail.com](mailto:jagmstar@gmail.com)<sup>3</sup>

**Abstract:** The article deals the task of comparing of agree of the results of expert evaluation conducted by several independent groups of experts of the software of teacher' rating of higher educational institutions. The resulting evaluation of the expert group calculated by the modified method of expert evaluation of software systems based on interval data analysis. The resulting evaluation of expert groups will can improve the "weak" aspects of the software system and can help conduct analyze the expert assessments too.

**Keywords:** expert evaluation, methods of interval data analysis, software system, teacher rating system.

## I. INTRODUCTION

One of the important tasks of the present is the research of one of the most important characteristics of software systems - quality. Under the quality of the software system, we understand the set of the software product properties, which characterizes its ability to meet the established or predicted needs of the customer, which he expressed in the form of user requirements for software in the early stages of its development [1]. One of the most important attributes of a software system's quality is the functional capability, that is, the ability of the system to perform its functions over a certain period of operation within predetermined limits and under certain operating conditions [2, 3]. Nowadays, a large number of leading scientists of Ukraine and the world is involved in the study of these issues and expert evaluation is a common method for estimating the quality of the software system. Expert (lat. Expertus - experienced) evaluation (expertise) - a method of obtaining summary information by the way of estimation a situation, event or phenomenon by a group of independent experts [4]. Such generalized information is obtained through an expert survey, which involves specialists in the area that interests the researcher. The task of an expert is to formulate his own opinion about the object under research or the phenomenon on a certain scale in accordance with the prescribed rules. The main problem of expert evaluation is the choice of competent in the study area experts, which have an unbiased attitude to the object of research and had a critical attitude towards all that evaluates, especially if it is a study of software systems whose lowly quality and may endanger the life of a person or of humanity [4].

In view of the above, the task of studying agree of expert evaluations of different groups in estimating software systems of any practical or theoretical direction is actuality.

## II. STATEMENT OF THE TASK

The evaluation of the software system quality begins from the early stages of development, including the definition of the specification requirements for the software system, analysis, design, etc. The quality estimation process at these stages usually involves design engineers who themselves evaluate some part of the system they are developing. After that, the results of their estimations are integrating and averaging. Of course, the resulting evaluation in this case will be overestimated [5].

Approaches and recommendations for the process of expert evaluation of the software systems quality in various scientific schools, studies by individual scientist cover various aspects of this issue. Some works try to formulate recommendations for the evaluation process. In the work [6] a critical analysis was made to cover all the attributes of the software systems quality that need to be accounting in the expert evaluation of systems. In particular, in the paper [7] we consider the methodology for evaluating the quality of web-projects. The authors proposed to evaluate quality not by formal numerical measurements, but in the form of relations and preferences with the application of the logic of antonyms [8]. The drawback of this method, in our opinion, is the using limit. It only can be used to obtain an end-user evaluation. In the works [9-11] a method of calculating the quality evaluation of software systems by a set of criteria from different groups of participants in the development process is proposed. The results of the expert evaluation of software systems quality according to various criteria by this method also allowed to developing a method of visualization of the information on the estimates using polar diagrams. However, this method has a number of remarks that discussed in the work [12].

One of the important drawbacks of the method is the exact formalization of the expert's opinion when evaluating a particular criterion [12]. The paper [6] provides information on the real economic and environmental impacts that may occur not only in the life of some company but also ordinary people, with insufficient evaluation of a particular project at various stages of its development. Therefore, the choice of experts and giving him the rights for objectively evaluate the proposed software systems is extremely actuality.

To take into account the possible questionable expert evaluations according to certain criteria, the method of expert evaluation of the software systems quality to be realization using the methods of interval data analysis are proposed [3,12-14].

Interval evaluation of the software system based on the set up of the upper and lower limits [14] of the permissible expert estimation according to a certain criterion:

$$x_{m,k} \in [x_{\min}; x_{\max}], \quad (1)$$

where  $x_{m,k} \in [1..10]$  - the evaluation is set by the expert on a certain criterion;  $m$  - the number of criteria for evaluation;  $k$  - the expert;  $x_{\min} = x_{m,k} - \delta \cdot x_{m,k}$ ;  $x_{\max} = x_{m,k}$ ;  $\delta$  - the percentage of deviation from evaluation the set by the expert, which can be determined for each project or expert, depending on the "degree of trust" to the expert.

The resulting expert evaluation has the form:

$$[X_k] = \sum_m [x_{\min}; x_{\max}] \cdot c_m, \quad (2)$$

where  $[X_k] = [X_k^-; X_k^+]$  - interval evaluation of the expert of the area, which takes into account the percentage of the expert's rating deviation from the nominal value;  $c_m$  - percentage indicator of importance of the criterion of software evaluation,  $\sum_m c_m = 1$ .

The resulting evaluation of the software system quality in this case will take the form:

$$[X] = \frac{\sum_k [X_k^-; X_k^+] \cdot q_k}{\sum_k q_k}, \quad (3)$$

where  $q_k$  - the weight indicator of the individual group;  $[X] = [X_{ex}^-; X_{ex}^+]$  - the resulting evaluation of all groups of experts, which is the interval of confidence to the software developing.

The condition of the agree of the obtained interval estimation of the software system quality to the admissible is [3,6]:

$$[X_{ex}^-; X_{ex}^+] \subset [X_{\min}; X_{\max}], \quad (4)$$

where  $[X_{\min}; X_{\max}]$  - the established interval of software evaluation, which is guaranteed to satisfy software developers.

### III. EXAMPLE OF THE EXPERT EVALUATION OF THE SOFTWARE OF TEACHER'S RATING OF HIGHER EDUCATIONAL INSTITUTIONS

In Fig. 1 schematically illustrated the process of expert evaluation the quality of the software system.

As we can see from the figure, the resulting evaluation of the software system quality depends on the estimates made by experts at the initial stage.

We will conduct a comparative analysis of evaluations put forward by different groups of experts on the example of the software of teacher' rating of higher educational institute. Shortly about the rating system of teachers. The first version of this system developed by the Master of the Department of Computer Science in 2017. In Fig. 2 shows the main window of the system.

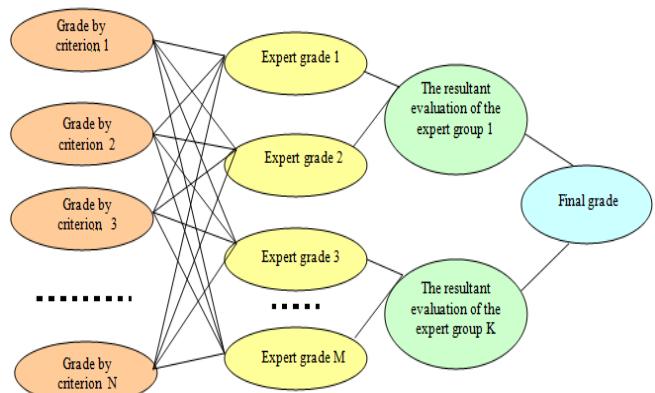


Fig. 1. The process of expert evaluation

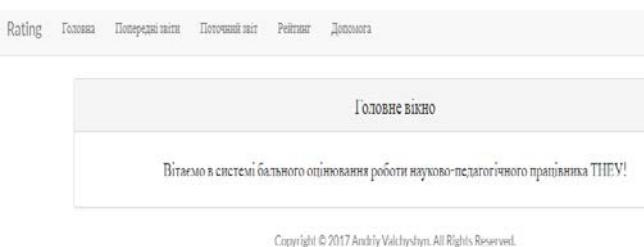


Fig. 2. The main window of the system of teacher' rating

On Fig. 3 is provide the general rating of university teachers.

| Загальний рейтинг |                                |         |
|-------------------|--------------------------------|---------|
| №                 | Ім'я                           | Бал     |
| 1                 | Шандук Сергій Костянтинович    | 843.75  |
| 2                 | Фурман Анатолій Васильович     | 834.25  |
| 3                 | Самченко Анатолій Олексійович  | 701.69  |
| 4                 | Лучко Михайло Романович        | 665.25  |
| 5                 | Савельєв Євген Васильович      | 576     |
| 6                 | Задорожний Ігор Вікторович     | 549.75  |
| 7                 | Дивак Микола Петрович          | 543.21  |
| 8                 | Дзиблік Олександр Валерійович  | 522.75  |
| 9                 | Ніколайчук Ярослав Миколайович | 518.875 |

Fig. 3. The general rating of university teachers

Evaluation of the work of the scientific and pedagogical workers of the university has a cumulative character, takes into account data for 5 years and it calculated by the following formula:

$$I = P + \sum_{i=1}^4 2^{-i} P_i, \quad (5)$$

where  $I$  - the evaluation of the work of the scientific and pedagogical workers for the last 5 years;  $P$  - the evaluation of the work of the employee for the reporting year;  $P_i$  - the evaluation of the work for the previous 4 years.

The work of the teacher evaluated in the context of such activities as research work, educational work, methodical work, organizational work, qualifications and additional criteria. On fig. 4 showed one of the windows for capability reporting.

| Звіт за 2017 рік   |   |   |   |   |   |   |
|--|---|---|---|---|---|---|
| <b>Види досягнутої кваліфікації</b>                            |   |   |   |   |   |   |
| <b>1. Наявність наукового ступеня</b>                          |   |   |   |   |   |   |
| <input type="checkbox"/> Доктор наук [30]                      | 9 | 9 | 8 | 9 | 8 | 8 |
| <input type="checkbox"/> Кандидат наук [20]                    | 9 | 9 | 8 | 9 | 9 | 8 |
| <b>2. Наявність вченого звання</b>                             |   |   |   |   |   |   |
| <input type="checkbox"/> Професора [20]                        | 8 | 8 | 9 | 9 | 7 | 9 |
| <input type="checkbox"/> Доцента [10]                          | 9 | 9 | 8 | 9 | 8 | 8 |
| <input type="checkbox"/> Старшого наукового співробітника [10] | 8 | 8 | 8 | 8 | 8 | 8 |
| <input type="checkbox"/> Провідний науковий співробітник [20]  | 9 | 9 | 9 | 9 | 8 | 9 |
| <input type="checkbox"/> Провідний науковий співробітник [20]  | 9 | 9 | 8 | 8 | 8 | 9 |

Fig. 4. Form to fill out the report by the teacher

To evaluate the quality of this software system in an objective way and the possibilities for its further perfection and improvement, it decided to conduct its evaluation. Leading departments of the Faculty of Computer Information Technologies were in the role of expert groups.

The software system evaluated according to the following criteria [12]:

1. The correctness of the work (the system must be isolated from external influences and the result of performing the functions should be correct in all conditions).
2. Protection from unauthorized access
3. Program reliability (the system must be resistant to various user-side influences)
4. Comfortable graphical user interface
5. Low cost of hardware resources (the system should not require high hardware costs of the computer)
6. Mobility (the system should have a small amount of memory, a small amount of processor time, etc., so that it can be used on any PC)
7. Scalability (improving the capabilities of the system by introducing a new functional)
8. Convenience of use
9. Speed
10. Completeness of functional requirements (the system must meet all of its functional requirements from the side of the subject area).

The software system was evaluated by the experts of the following categories: expert in the area (in the table - EA), business analyst (BA), software architect (SA) and expert of user interface (EI). Percentage coefficient importance of the criterion  $c_m$  for each of the criteria listed above, accordingly 0,05; 0,05; 0,05; 0,1; 0,05; 0,05; 0,05; 0,1; 0,1; 0,4. System developers want to achieve system quality with a minimum threshold of 80%. Below in the tables 1-3 shows the upper limit for the expert's estimate. The lower limit of the evaluation will be formed with a deviation 5% from the upper

one in order to take into account the "doubts" of experts of the setting score.

TABLE 1. EVALUATION OF THE SOFTWARE SYSTEM BY EXPERTS FROM THE DEPARTMENT 1

| №  | EA         | BA         | SA          | SA          | EI          | EI |
|--|------------|------------|-------------|-------------|-------------|----|
| <b>1</b>                                       | 9          | 9          | 8           | 9           | 8           | 8  |
| <b>2</b>                                       | 9          | 9          | 8           | 7           | 8           | 7  |
| <b>3</b>                                       | 9          | 8          | 9           | 9           | 9           | 8  |
| <b>4</b>                                       | 8          | 8          | 9           | 9           | 7           | 9  |
| <b>5</b>                                       | 9          | 9          | 8           | 9           | 8           | 8  |
| <b>6</b>                                       | 8          | 7          | 9           | 9           | 9           | 8  |
| <b>7</b>                                       | 8          | 8          | 8           | 8           | 8           | 8  |
| <b>8</b>                                       | 9          | 9          | 9           | 9           | 8           | 9  |
| <b>9</b>                                       | 9          | 8          | 9           | 7           | 9           | 9  |
| <b>10</b>                                      | 9          | 8          | 8           | 8           | 8           | 9  |
| Weight indicator of each expert                | 7          | 8          | 9           | 9           | 9           | 9  |
| The resulting evaluation of a group of experts | [8,35;8,8] | [7,79;8,2] | [7,91;8,33] | [7,96;8,38] |             |    |
| The resulting evaluation of the project        |            |            |             |             | [7,97;8,39] |    |

TABLE 2. EVALUATION OF THE SOFTWARE SYSTEM BY EXPERTS FROM THE DEPARTMENT 2

| №  | EA         | BA        | SA        | EI        |
|--|------------|-----------|-----------|-----------|
| <b>1</b>                                       | 7          | 8         | 8         | 7         |
| <b>2</b>                                       | 8          | 9         | 7         | 8         |
| <b>3</b>                                       | 8          | 8         | 7         | 8         |
| <b>4</b>                                       | 7          | 8         | 8         | 7         |
| <b>5</b>                                       | 7          | 9         | 7         | 7         |
| <b>6</b>                                       | 7          | 9         | 7         | 7         |
| <b>7</b>                                       | 7          | 9         | 8         | 7         |
| <b>8</b>                                       | 7          | 7         | 7         | 8         |
| <b>9</b>                                       | 8          | 8         | 7         | 7         |
| <b>10</b>                                      | 7          | 8         | 7         | 7         |
| Weight indicator of each expert                | 7          | 8         | 9         | 9         |
| The resulting evaluation of a group of experts | [6,84;7,2] | [7,7;8,1] | [6,8;7,2] | [6,8;7,2] |
| The resulting evaluation of the project        |            |           |           | [7,1;7,4] |

TABLE 3. EVALUATION OF THE SOFTWARE SYSTEM BY EXPERTS FROM THE DEPARTMENT 3

| №        | BA | EA | EA | EA | SA | SA | SA | EI | EI | EI |
|----------|----|----|----|----|----|----|----|----|----|----|
| <b>1</b> | 9  | 10 | 8  | 8  | 9  | 9  | 8  | 7  | 9  | 9  |
| <b>2</b> | 8  | 9  | 7  | 4  | 8  | 3  | 7  | 9  | 7  | 8  |
| <b>3</b> | 8  | 9  | 5  | 4  | 8  | 5  | 8  | 7  | 7  | 8  |
| <b>4</b> | 8  | 7  | 7  | 4  | 9  | 8  | 6  | 7  | 9  | 8  |
| <b>5</b> | 9  | 10 | 10 | 8  | 9  | 8  | 8  | 8  | 9  | 9  |
| <b>6</b> | 7  | 10 | 8  | 4  | 7  | 8  | 6  | 4  | 6  | 7  |
| <b>7</b> | 8  | 10 | 10 | 7  | 8  | 9  | 7  | 5  | 8  | 8  |

|  |           |           |    |           |           |           |   |   |   |   |
|--|-----------|-----------|----|-----------|-----------|-----------|---|---|---|---|
| <b>8</b>                                       | 8         | 9         | 8  | 7         | 9         | 8         | 7 | 6 | 7 | 8 |
| <b>9</b>                                       | 8         | 10        | 10 | 8         | 8         | 7         | 8 | 8 | 6 | 7 |
| <b>10</b>                                      | 8         | 10        | 7  | 7         | 7         | 9         | 8 | 7 | 6 | 7 |
| The resulting evaluation of a group of experts | [7,6;8,1] | [7,5;7,9] |    | [7,4;7,8] |           | [6,8;7,1] |   |   |   |   |
| The resulting evaluation of the project        |           |           |    |           | [7,2;7,6] |           |   |   |   |   |

Consequently, from the expert evaluation of the rating system of the university teachers, the main quantitative estimates of which given in the tables, we can conclude that in fact the results of the evaluation of only one of the three expert groups are consistent with the initially established interval of evaluation of the software system, which satisfies the developers.

#### IV. CONCLUSION

The paper is devoted to the problem of studying and evaluating the quality of software systems. The methods of expert evaluation are considered. It indicated that most of them carry only theoretical recommendations for improving the evaluation process. Some methods aimed at improving the visual presentation of the evaluation results. The results of the analysis showed the actuality and importance of paying special attention to the selection of experts who would evaluate software systems so that their opinion was objective and unbiased. It proved that, in order to avoid "doubts" regarding the evaluation of a particular criterion, use the methods of interval data analysis. The method of expert evaluation of software systems based on the analysis of interval data is developed, the result of which is to check the agree of the resulting interval estimation according to the project proposed by an independent group of experts and the set the interval of evaluation of the software system that satisfies the developers. On the example of the expert evaluation of the rating system of the university teachers, the diversity of opinions of different groups of experts shown.

#### REFERENCES

- [1] Pomorova, O.V., Hovorushchenko, T.O, "Modern problems of software quality assessment", *Radioelektronni I kompiuterni sistemy*, no.5, pp.319-327, 2013.
- [2] International Standard ISO/IEC 9126. Information technology – Software product evaluation – Quality characteristics and guidelines for their use. *International Organization for Standardization International Electrotechnical Commission*, Geneva, 1991.
- [3] S. Krepych, A. Dyvak, M. Dyvak, I. Spivak, "The method of providing of functional suitability of elements of the device of formation of signal in electrophysiological way of classification tissues surgical wound", *13<sup>th</sup> International Conference Perspective Technologies and Methods in MEMS Design, MEMSTECH 2017 Proceedings*, pp.183-186, 2017.
- [4] I. Galyan, "Psychodiagnostics: Textbook," Kyiv Academic Edition, p. 464, 2011.
- [5] Y. Ryabokon, "Software cost estimation", *Electrical and Automation system*, vol. 1(82), pp.117-124, 2015.
- [6] Grytsyuk Yu., Gritsyuk P., "Modern problems of scientific evaluation of the applied software quality", *Scientific Bulletin of the NLTU of Ukraine "Information Technologies and Modeling in Economics"*, №. 25/7, pp. 284-294, 2015
- [7] Berko, A., Alekseeva, K., "Estimation of the information resources quality in WEB-projects", *"Actual problems of economy"*, №10 (136), pp.226-234, 2012.
- [8] Golota Y., Tysenko V., Falkov D., "The logic of antonyms is the theoretical basis for the formation of complex assessments based on expert estimates of individual parameters," *Modeling of intellectual processes of design and production: Materials II internationally scientific-practical conference*, Minsk, p.166-167, 1998.
- [9] I. Morgan, "Method of expert evaluation of software quality", *Materials of the International Scientific and Practical Conference of Postgraduate Students and Students "Software Engineering 2011"*, vol.2(6), pp.117-124, 2011.
- [10] Morgan I., Botsula M., "New method and information technology for data processing for quality management of electronic training courses", *International scientific and technical magazine "Information technologies and computer engineering"*, №3, pp.25- 33, 2014.
- [11] Grytsyuk Yu., Buchkovskaya A., "Visualization of the results of expert evaluation of software quality using polar diagrams", *Scientific Bulletin of NLTU of Ukraine*, Vol.27, No.10, pp.137- 145, 2017.
- [12] I.Spivak, S.Krepych, S. Budenchuk, "Methods and means of expert evaluation of software systems on the basis of interval data analysis", *14<sup>th</sup> International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering* pp.101-127, 2018
- [13] S. Krepych, I.Spivak, "Estimation of the time complexity of the Monte Carlo method and interval analysis of data for determining the functional suitability of REC", *Modern Computer Information Technologies: Materials of the Third All-Ukrainian School-Seminar for Young Scientists and Students ACIT'2013*, Ternopil, pp.36-37, 2013.
- [14] I. Spivak, M. Dyvak, "Tolerance estimation of the parameters of "input-output" dynamic model on the basis of interval data analysis", *Proceeding of International Conference CADSM'2005*, Lviv-Polyana, pp.151-153, 2005.

# Software for Modelling the Air Pollution by Vehicles

Vasyl Tymchyshyn, Natalia Porplytsya, Andriy Melnyk, Bohdan Tymchyshyn

Department of Computer Science, Ternopil National Economic University, UKRAINE, Ternopil, 8 Chekhova str.,  
email: [tymchyshyn@gmail.com](mailto:tymchyshyn@gmail.com), [ocheretnyuk.n@gmail.com](mailto:ocheretnyuk.n@gmail.com), [melnik.andriy@gmail.com](mailto:melnik.andriy@gmail.com), [tymchyshyn.bohdan@gmail.com](mailto:tymchyshyn.bohdan@gmail.com)

**Abstract:** In this work the issues of the ecological safety of motor transport, the impact of motor vehicles on the environment are highlighted. Directions and measures concerning the raised ecological safety of motor transport are determined.

**Keywords:** motor vehicles, emissions from motor vehicles, automobile fuel, software complex.

## I. INTRODUCTION

One of the most powerful sources of urban air pollution is road transport, the increase in the number of which has led to a sharp deterioration of sanitary conditions in large cities. Motor transport pollutes the air with toxic components exhaust gases, fuel pairs, tire wear products, brake linings, creates noise and vibration. Emissions from motor vehicles negatively affect the physiological state of humans and animals, pollute water, destroy soils, vegetation, building materials, architectural and sculptural monuments, cause corrosion of metals, etc [1].

The problem of a comprehensive solution of the ecological problems in the city consists in the absence of a unified monitoring system, one of the fundamental principles of which is the interconnected network of observation, control, collection and processing information for the analysis, assessment and forecasting the state of environmental pollution. Therefore, the task of realization of the software complex for automation of the process of monitoring and visualization atmosphere pollution by harmful emissions of vehicles (in particular, nitrogen dioxide) in time is actual.

## II. METHODS FOR MEASURING THE CONCENTRATIONS OF HARMFUL SUBSTANCES IN THE NEAR-GROUND LAYER OF THE ATMOSPHERE

There are two main methods of measuring the harmful substances in the air: chemical analysis and microbiological analysis.

The chemical analysis of air provides information on the qualitative and quantitative composition on the basis of which it is possible to predict the degree pollution and plan the implementation of measures to control air quality. Detects indicators such as dust, sulfur dioxide, nitrogen dioxide, carbon monoxide, phenol, ammonia, hydrogen chloride, formaldehyde, benzene, toluene, etc. This technique allows to determine the presence in the air of volatile organic compounds with a boiling point of 40 to 250 ° C, affecting human health (phenols, phthalates, organic acids, aromatic compounds, ethers, morphine and other compounds - up to 250).

Microbiological analysis air allows us to establish the presence of biological aerosols (bacteria and fungi). It is

necessary to conduct detection pathogenic microorganisms according to such indicators as: total number of microorganisms, gold staphylococci, mold and yeast. Gas analysis of air is carried out using a device called a gas analyzer. Gas analyzer is a measuring device for determining the qualitative and quantitative composition of gas mixtures.

Depending on the pressure in the reaction chamber, gas analyzers of atmospheric and low pressure are distinguished. Gas analyzers with built-in NO<sub>2</sub> / NO converters produce analytical signals for NO, NO<sub>x</sub> and NO<sub>2</sub> simultaneously or sequentially. Table 1 gives a comparative characteristics NO<sub>2</sub> measuring devices.

TABLE 1. COMPARATIVE CHARACTERISTICS NO<sub>2</sub> MEASURING DEVICES

| Name              | DGS-NO <sub>2</sub><br>968-037 | Ug-2                    | Polar            |
|-------------------|--------------------------------|-------------------------|------------------|
| Price(UAH)        | 2100                           | 4000                    | 6000             |
| Dimensions        | 44.5 x 20.8<br>x 8.9 mm        | 110 x<br>105x<br>200 mm | 148×164×80<br>mm |
| Measurement error | 15%                            | 25%                     | 20%              |

Nitrogen dioxide is a toxic substance, which is why the important task is to control the release this compound.

Many devices have been developed to measure nitrogen dioxide, but one of the most effective uses is the use of the DGS-NO<sub>2</sub> 968-037 sensor as it many advantages, such as low price, high resolution etc.

Fig. 1 shows the general view of the sensor DGS-NO<sub>2</sub> 968-037.

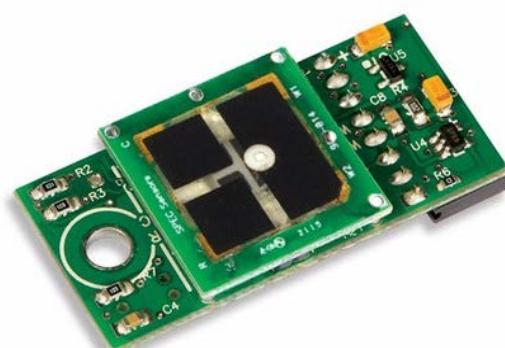


Fig.1. General view of the sensor DGS-NO<sub>2</sub> 968-037



Fig.2. Connected DGS-NO<sub>2</sub> 968-037 to USB-Bridge UART

Note, that the software and hardware complex for modeling the pollution of the atmosphere by motor transport, designed within this work, it contains a sensor DGS-NO<sub>2</sub> 968-037 for the collection experimental data.

### III. FEATURES OF COLLECTING THE EXPERIMENTAL DATA

To collect experimental data on the pollution of air by harmful emissions of vehicles, the study was conducted at the crossroads of the streets Za Rudkoyu str. and Chekhova srt., because precisely at this point one of the most intense automobile streams of the "New World" micro district in Ternopil.

Experiment date: 27.10.17, time from 14.00 to 15.00, air temperature 11°C, air humidity 74%. For effective measurement nitrogen dioxide concentration, the sensor should be located at the point of the road as close as possible to the asphalt surface, because in this point the concentration of NO<sub>2</sub> is maximal.

Table 2 shows the averaged NO<sub>2</sub> concentrations obtained over an hour by measuring the concentration of the DGS-NO<sub>2</sub> 968-037 sensor. The format of the output is: Sensor number [XXXXXXXXXXXX], PPB (Part per billion) [0: 999999], Temperature (°C) [-99:99], RH [0:99], RawSensor [ADCCount], TempDigital, RHDigital, Day [0:99], Hour [0:23], Minute [0:59], Second [0:59].

TABLE 2. AVERAGED NO<sub>2</sub> CONCENTRATIONS

| Time<br>(min) | NO <sub>2</sub> , ppb | NO <sub>2</sub> , ppm | NO <sub>2</sub><br>mg/m <sup>3</sup> |
|---------------|-----------------------|-----------------------|--------------------------------------|
| 10            | 20,9                  | 0,0209                | 0,0423                               |
| 20            | 11,15                 | 0,01115               | 0,02259                              |
| 30            | 8,683333              | 0,008683333           | 0,01759                              |
| 40            | 14,76667              | 0,014766667           | 0,02992                              |
| 50            | 9,5                   | 0,0095                | 0,0192                               |
| 60            | 3                     | 0,003                 | 0,00608                              |

Note, that the measurements were made at a frequency of 10 seconds, so Figure 3 shows the averaging values of the concentration of nitrogen dioxide in 10 minutes.

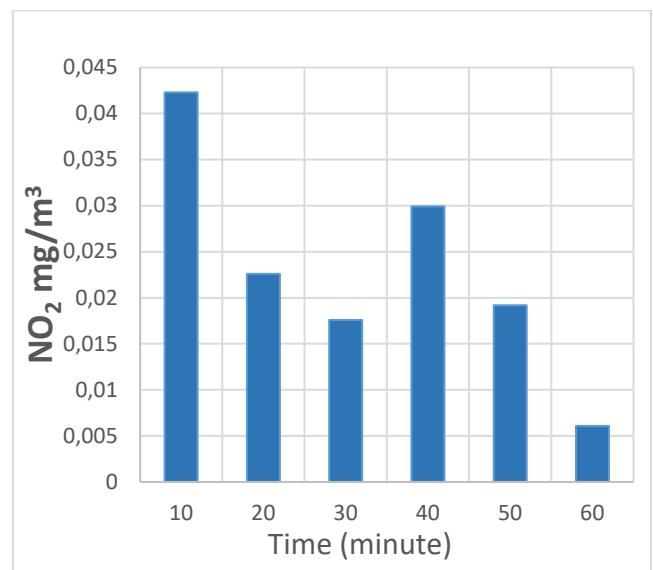


Fig.3. The fragment of the measured values of the nitrogen dioxide concentration

### IV. MATHEMATICAL MODEL FOR MODELLING THE POLLUTION OF THE ATMOSPHERE BY HARMFUL EMISSIONS FROM VEHICLES

Regular measurement of atmospheric pollution by harmful emissions vehicles and establishment actual concentrations pollution requires a huge amount resources, in particular gas analyzers, special equipment for ecological monitoring systems, which is unrealistic even for countries with high economic levels and social development. An alternative to this is the development monitoring systems with mathematical models in their composition, which, based on a limited sample data, make it possible to predict time changes the concentrations harmful emissions. In this case, it is expedient to use interval discrete dynamic models, and to identify them based on the analysis of interval data, as shown in such works [2-5].

In order to construct the model, is necessary a set of experimental data. Within the scope of work, a mobile measurement unit, based on a personal computer and a digital gas sensor type DGS-NO<sub>2</sub> 968-037, was used for their obtaining. Measurement was carried out at the crossroads of the streets Ruska str. - Zamkova str, in Ternopil city every second. In order to compensate for a random measurement error, which is normally distributed with zero mathematical expectation, the measured instantaneous values were averaged in a window with a duration of 20 minutes. The fragment measured values of the concentration of nitrogen dioxide, temperature, humidity and traffic intensity of cars, at the crossroads of the streets Ruska-Zamkova, Ternopil, with a discrete of 20 minutes, is given in Table 3.

TABLE 3. THE OUTPUT DATA FOR MODELING THE DYNAMIC OF NO<sub>2</sub> CONCENTRATIONS

| Time | v <sub>-k</sub> | v <sub>+k</sub> | Tepm | Rh     | Cars |
|------|-----------------|-----------------|------|--------|------|
| 0:20 | 0,0336          | 0,04545         | 4    | 73,534 | 168  |
| 0:40 | 0,0232          | 0,03147         | 4    | 73,834 | 152  |

TABLE 3 (CONTINUE)

|       |         |         |     |         |     |
|-------|---------|---------|-----|---------|-----|
| 1:00  | 0,01782 | 0,02411 | 4   | 73,6424 | 147 |
| ...   | ...     | ...     | ... | ...     | ... |
| 12:00 | 0,06925 | 0,09369 | 5   | 94,2149 | 711 |
| 12:20 | 0,07337 | 0,09927 | 5   | 94,3498 | 694 |
| 12:40 | 0,07474 | 0,10112 | 4   | 96,3429 | 681 |
| 13:00 | 0,07394 | 0,10004 | 4   | 95,8456 | 779 |
| ...   | ...     | ...     | ... | ...     | ... |
| 23:00 | 0,04608 | 0,06235 | 2   | 77,4495 | 254 |
| 23:20 | 0,04486 | 0,06069 | 2   | 77,9879 | 219 |
| 23:40 | 0,04478 | 0,06059 | 2   | 78,3345 | 196 |
| 0:00  | 0,04419 | 0,05979 | 2   | 78,1432 | 172 |

Further, a well-known structural identification method, built on the basis behavioral models of the bee colony [6, 7], was used to construct a mathematical model for predicting the concentrations harmful emissions [8, 9].

The using of the method involves the transformation of structures of interval discrete models by operators  $P(\Lambda_{mcn}, F)$ ,  $P_\delta(\Lambda_{mcn}, F)$ ,  $P_N(F, I_{\min}, I_{\max})$  and through holding selection procedures by operators  $D_1(\lambda_s, \lambda'_s)$ ,  $D_2(\lambda_s, \Lambda'_s)$  in order to provide the reduction on each iteration of the goal function values for optimization task of structural identification the interval discrete dynamic model[8].

As a result of using the structural identification method, an adequate mathematical model for predicting the concentrations of harmful emissions from vehicles was obtained:

$$\begin{aligned} \hat{v}_k = & 0.0365 + 0.3541 \cdot \hat{v}_{k-1} \\ & + 0.118 \cdot \hat{v}_{k-1} \cdot \hat{v}_{k-3} + 0.5059 \cdot \hat{v}_{k-1} \cdot u_{3,k-1} / u_{3,k-1} - \\ & - 0.01544 \cdot \hat{v}_{k-2} \cdot u_{3,k-1} / u_{3,k+1}, \end{aligned} \quad (1)$$

where  $k=4\dots72$ ;  $\hat{v}_k$  – the predicted concentration nitrogen dioxide value in k moment of time;  $\vec{u}_3 = (u_{3,0}, \dots, u_{3,k+1})$  – known input variables vector (the intensity of traffic flows).

Note, the concentrations of measured values of harmful emissions NO<sub>2</sub> (at points  $k = 0 \dots 3$ ) should be set as initial conditions in the interval  $\pm 0,5\%$ , for the modeling with using linear discrete equation (1).

As we see, the obtained mathematical model reflects the dynamics concentrations of nitrogen dioxide, with a discrete time value 20 minutes. To use it, it is sufficient to set initial values of the measured concentrations, the temperature and humidity forecast, which is not a problem at short time intervals (for example, in the interval of one day). Note that the finded model can be used to model the concentrations harmful emissions in other city dots, provided that the parameters of the model are clarified.

## V. DESIGNING OF THE MONITORING SYSTEM'S ARCHITECTURE

Figure 4 shows the developed architecture of the software system for automation monitoring and visualization pollution the near-ground layer atmosphere by harmful emissions of

vehicles. As can be seen from the figure in this typical architecture, there are three logical layers [10]:

- 1) user interface (visualization layer);
- 2) data processing (business logic layer);
- 3) data access layer.

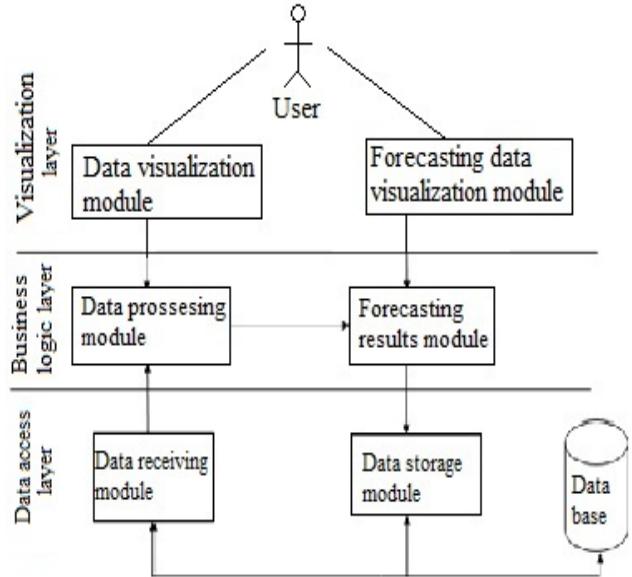


Fig.4. Architecture of the software system for modeling the air pollution by vehicles

At the first layer, the modules with which the user works and which are intended to visualize the results of the research are presented. This level does not have a direct connection with the database and the main business logic, in terms of security.

At the second layer, all data processing is carried out. This level is represented by the following modules:

- data processing module;
- forecasting results module – the main system module, which, implements the process of forecasting the concentration harmful emissions vehicles in a specific city point based on the model (1).

At the data access level, modules are stored, through which the business logic level interacts with the database using CRUD operations.

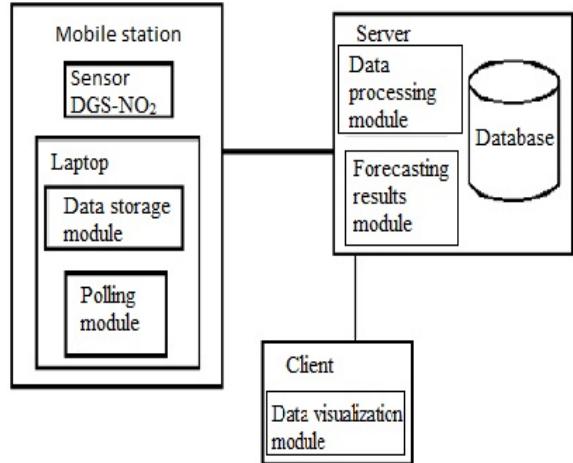


Fig.5. System modules placement

Figure 5 shows the system modules placement for modeling atmosphere pollution by vehicle. As can be seen from the figure, the modules are located on different hardware.

A mobile station measuring NO<sub>2</sub> level is based on a notebook with a Windows operating system and a sensor called "SPEC Sensors DGS-NO<sub>2</sub> 968-037". The DGS-NO<sub>2</sub> 968-037 sensor is equipped with an UART-to-USB adapter, which allows you to connect it to your computer through the USB interface. For the correct sensor operation, it should be installed CP210x USB to UART Bridge VCP driver and terminal TeraTerm, on the laptop.

The sensor connection is carried out similarly to the previous case - via the USB interface. Measurement of instant concentrations of NO<sub>2</sub> is carried out every second. Data is recorded in a log file and transmitted to the server using a Wi-Fi connection.

The monitoring station is the server where the NO<sub>2</sub> measured concentrations are located, software for constructing mathematical models to forecast the time distribution of the indicated concentrations.

Also, the monitoring station implemented a server part web-based system to display the modeling results and archive data on the level concentrations harmful emissions vehicles in the atmosphere city Ternopil.

On the user's side, deployed a client-side web-based system that lets you monitor real-time emissions of nitrogen dioxide into the air.

To use the website, the user will need to access the Internet from the computer that is used, and any web browser that supports HTML5 and CSS3 standards is installed.

Figure 6 shows the look of the home page of the website.



Fig.6. The website main page

As shown in Figure 6, in order to allow the user to view the concentration data dioxide, he must select from the dropdown list the control point (street crossings), after which a concentration graph of the daily cycle of NO<sub>2</sub> will be displayed before it.

## VI. CONCLUSION

The paper considers an approach for modeling a daily cycle of changes in nitrogen dioxide concentrations within a road single section. A method is developed for operative and automated obtaining of experimental data. Designed and developed software architecture for modeling the atmosphere pollution by harmful emissions of vehicles. The proposed method approbation for the receipt and processing experimental data on NO<sub>2</sub> concentrations, as well as software developed on the example of modeling the distribution harmful emissions in the city of Ternopil.

## REFERENCES

- [1] Vasyukova G.T. "Ecology": a textbook for students, Kyiv: Condor, 2009, 311 p. (in Ukrainian)
- [2] Dyvak M. "Tasks of mathematical modeling the static systems with interval". Ternopil, Ukraine: Ekonomichna dumka, 2011, 216 p. (in Ukrainian)
- [3] Ivakhnenko A.G., "Long-term forecasting and management of complex systems". Kyiv, Ukraine: Tekhnika, 1975, 311 p. (in Russian)
- [4] T. Dyvak, "Parametric identification of interval difference operator on the example of micromodel for distribution of humidity in the drywall sheets in the process of drying", *Information Technologies and Computer Engineering: international Scientific Journal*, vol. 3, pp. 79-85, 2012. (in Ukrainian)
- [5] N. Porplytsya, M. Dyvak, "Interval difference operator for the task of identification recurrent laryngeal nerve", *Computational Problems of Electrical Engineering: Proceedings of the 16th International Conference (CPEE)*, pp. 156-158, 2015.
- [6] D. Karaboga and B. Basturk, "A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm", *Journal of Global Optimization*, vol. 39, pp. 459-471, 2007.
- [7] Karaboga D., Gorkemli B., Ozturk C., Karaboga N. "A comprehensive survey: artificial bee colony (ABC) algorithm and applications", *Artificial Intelligence Review*, 42(1), pp. 21-57, 2014.
- [8] N. Porplytsya and M. Dyvak, "Synthesis of structure of interval difference operator using artifitional bee colony algorithm", *Inductive modeling of complex systems*, vol. 5, pp. 256-269, Kyiv, Ukraine, 2013. (in Ukrainian)
- [9] M. Dyvak, N. Porplytsya, I. Borivets, M. Shynkaryk "Improving the computational implementation of the parametric identification method for interval discrete dynamic models", *12th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT)*, pp. 533-536, 2017.
- [10] W. Eckerson, "Three Tier Client/Server Architecture Achieving Scalability, Performance, and Efficiency in Client Server Applications." *Open Information Systems* vol, 87, no. 3, pp. 330-333, 2007.

# Next-generation Serverless System for Contextual Search Based on Rich Media Content

Vladyslav Holubiev<sup>1</sup>, Bohdan Ihnatiuk<sup>2</sup>, Iryna Voytyuk<sup>3</sup>

Faculty of Computer Information Technologies, Ternopil National Economic University, UKRAINE, Ternopil, 8 Chekhova str., email:  
[vladyslav.holubiev@gmail.com](mailto:vladyslav.holubiev@gmail.com)<sup>1</sup>, [bohdan.ihnatiuk@eleks.com](mailto:bohdan.ihnatiuk@eleks.com)<sup>2</sup>, [i.voytyuk@tneu.edu.ua](mailto:i.voytyuk@tneu.edu.ua)<sup>3</sup>

**Abstract:** The purpose of this research is building a next-generation knowledge management system based on the most recent developments in the industry of serverless computing. Such approach reimagines a conventional way people build software by abstracting away the complexity of maintaining highly available, highly scalable systems. The software combines many ways to parse rich media and text content such as image, audio, video to build a contextual search index.

**Keywords:** serverless system, contextual search, rich media content.

## I. INTRODUCTION

One of the problems which pop up on a day to day job of office workers - is facing a huge amount of information. Even more adds up to the problem when these files are scattered across different cloud systems, storage providers and there is no single place to store all. The next problem is that these cloud-based systems have a very inconsistent user interface and the average office employee should spend hours and hours learning a new UI, which drastically reduces performance and prolongs onboarding time. After some quick calculations, it turns out this process takes up to 500 hours, which causes new problems and diseases that emerged only in the 21st century - information overload and fatigue from decision making. As a result, an office worker cannot make critical decisions and loses productivity.

The next problem of the existing cloud storage providers - they have limited search interface. No single solution on a market can do powerful analysis and extract all the information from files and parse the content. Each of them focuses on one aspect of the user's field of knowledge, while today the organization can operate with an extremely diverse range of file formats. Even the largest search engines are focused on a thorough analysis of web pages while closing eyes on office workers in enterprise companies.

Talking about enterprises with thousands of employees, the amount of wasted money measured in millions. So it is critically important to stick to the discipline and the common idea of knowledge sharing in the organization in order not to be distracted by the search for information. Following some simple practices as avoiding people known as knowledge holders, making knowledge easy to find and share will bring you a long way ahead.

In addition, it is necessary to take into account achievements of the 21st century in the field of machine learning. This allows you to use new areas for analyzing files for a self-regulating organization. With a large number of

files, human assistance is no longer effective and error-prone.

After combining all of the above-mentioned with the latest technologies of parallel computing, the system was born, which solves the problem of overloading machine resources. New developments in cloud tech and so-called "serverless computing" make it possible to use resources extremely cost-effectively, paying only for the time spent by your code per second of time and gigabyte of memory [1]. A new unit of measurement was born - GB/s.

Per-second payment cycle saves the organization's budget since computer hardware does not work at nights and other hours that are not considered to be working in the office world. Instead, the company rents a CPU power of another company which owns loads of efficient hardware. Thanks to global coverage, the tenant company provides 24-hour service in all time zones, thereby minimizing costs by reducing latency. Even by putting 100 data centers around the world in the largest cities, gives a coverage which ensures <50 ms latency regardless of your geolocation.

After analyzing the problem, assessing available solutions and researching innovative technology - a new knowledge management system was born which steps up the game.

## II. TECHNOLOGIES BEHIND CONTENT CATEGORIZATION AND SEARCH INDEX

The process of synchronization described in this paper is aimed at reducing the negative effects outlined above. The main point is assuming file synchronization happens in terms of some folders structure. Given a normal distribution of files per folder, we can apply parallelism at the hierarchy level. With a folder structure, you can divide it into identical shards in order to squeeze the most performance of available computing resources.

In this case, several components that put a risk on files are fixed at once. Firstly, there is no queue now. Due to an efficient allocation of a tree structure of a folder hierarchy, it became possible to split all files into the so-called "branches" and process them separately and independently. Because each branch has the same small size, it takes only a few seconds to import, so the chance that there will be some kind of error at the moment is diminishingly small. This solves the problem of importing through a large number of files. The second improvement is that instead of using one server, there are now a large number of servers that perform the same job. At first glance, it may seem that it will be less effective since organizations will have to pay for all these servers and maintain them in proper condition, which multiplies the resources spent on one server. Instead, the developed

software system is based on the technology of serverless computing, which can be used to leverage computing power from other companies. This is a new form of business that has gained popularity.

Over the last 2017, serverless computing has gained a significant growth. Thus, organizations of different sizes around the world appreciate benefits of hassle-free computing power without wasting their resources to support many own servers, which stand idle at night. Just for granted, at your will, you get essentially "unlimited" number of parallel computations, which do not exceed five minutes each. Let's consider a way of processing text information.

So from the previous step, the system received input files in different custom formats. To search for these files you need to convert them to text format. Typically, several modern file processing techniques are used. Let's consider in detail the technique of serverless computing. At the first step, let's collect a sample of all the different formats that can be encountered when an import is finished - a representative sample of typical files that the office worker uses most often. Next, consider the methods of processing each of these file types. Often, text files carry a lot of unimportant information to look for. Therefore, the system will process them for cleaning out irrelevant characters. This process is called "tokenization". It also includes the process of bringing words to its original form. After this stage of processing, we have text in the output that is suitable for transferring to the search index. The same principles then apply to all other types of files after they pass the step of extracting information.

After analyzing file types that are used by office workers, the most common types of photo materials were found. In 70% of cases, these are scanned documents. The rest of the files are general-purpose photographs, where no text is depicted, but objects are present (more on this later).

Let's choose two tactics of image processing (Fig. 1).

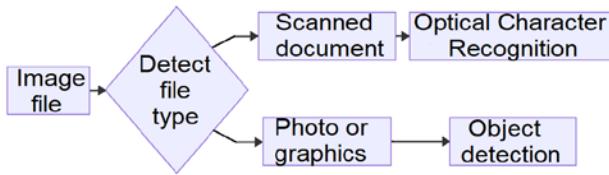


Fig. 1. Two tactics of image processing

The first tactic involves the use of machine learning algorithms that were trained on millions of characters with different projections and distortions. The software for this task is open source and is not copyrighted. The knowledge management software system has a built-in neural network model for recognition documents in three most popular languages such as English, French, and German [2].

For example, let's take a photo of a Ukrainian bachelor's diploma. As shown in Fig. 2, a trained model recognized the text in the image with close to 100% accuracy. An office worker can find this document by quoting a word or a phrase from an image. Such rich media integration greatly expands capabilities of conventional search engines.

In case the model was unable to extract enough text from the image file after the first step, it means that we are dealing with the type of photo file without text. In such situation, another model of the neural network, which is trained on

millions of images, will be used. Usually, such networks are created by training them by hand using a supervised training method. Unlike to unsupervised training, a human being required to teach the model range of objects. During this process, the model makes links between elementary image properties and the vector of available objects for recognition.



Fig. 2. Example of Optical Character Recognition

However, it's worthwhile to leverage readily available developments of Google company, such as TensorFlow (open source license). This software processes an input image and outputs a scalar vector of key phrases depicted in the photo. This example is illustrated in Fig. 3.

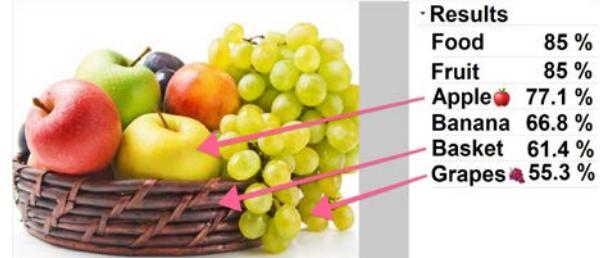


Fig. 3. Example of recognition the scene and objects in the photo

Given an image of basket with fruit, TensorFlow model outputs a list of objects.

Given an image of a basket, TensorFlow model outputs a list of fruit, such as apple, grapes, and the basket itself. Imagine a company file storage, filled with thousands of photos which are searchable only by file name. This kind of rich media experience opens new ways for content findability. Furthermore, detectable objects are limited to a list of known vocabulary words, so it makes easy applying search faceting to the list of objects. In a couple of clicks, an employee can drill down thousands of search results in orders of magnitude.

It is worth mentioning the fact that each recognized item is followed by a percentage of confidence, which will then allow you to weed out variants that are not relevant.

A video of the 136th episode of TNEU "Kaleidoscope of Events" was taken to demonstrate possibilities of video analysis for object detection in video media. The architecture of the video recognition process is shown in Fig. 4. The following recognition categories are available: people, faces, logos, celebrities.

The theory behind video processing is pretty much the same as images, but the scale of processing power required grows exponentially [3]. The system needs plenty of CPU

resources to analyze each frame with the same grade of detail as each photo. That's where serverless computing shines in its glory. For a case when we use our own trained model for detection we just rent out a couple of seconds of runtime in cloud providers and run the analysis there. As another option, cloud providers do offer a dedicated service for video recognition. From a viewpoint of knowledge management system, such setup is considered serverless as well, as it does not provision any servers for video analysis at all.

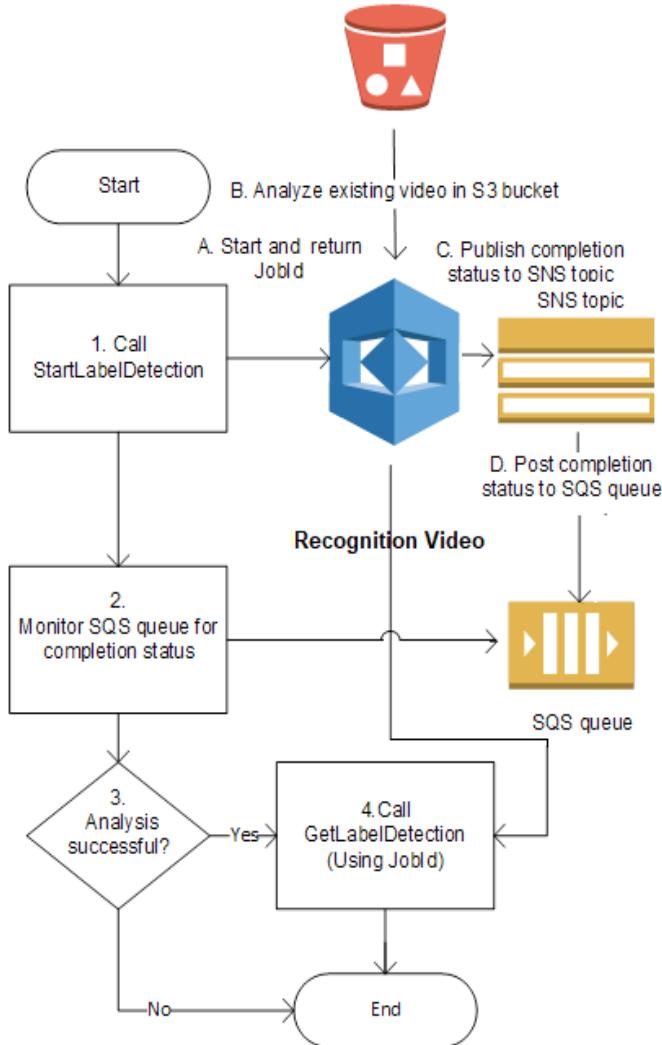


Fig. 4. The architecture of the video file analysis process

Video processing is a very complicated process and requires enormous computing resources. Therefore, in this case, the software system solves the problem by using the AWS Rekognition software solution for analyzing video files. It's enough to send the link to the downloadable file from the cloud storage and check the status of the recognition process. As a result of analysis of the latest episode of the TNEU "Kaleidoscope of Events" it discovered plenty of entities which appeared in a frame, as shown in Fig. 5.

After manual checking the outcome of analysis on my own and watching the video, it is safe to come to the conclusion that the confidence score of the keywords is acceptable and it's good to go in the knowledge management system. Similar to images, this opens a whole new set of ways for employees to find company's content scattered across many videos. This

means less time for people to put text descriptions to videos just to make them searchable.

#### ▼ Objects and activities

|               |          |        |           |                  |
|---------------|----------|--------|-----------|------------------|
| Human         | Person   | People | Military  | Military Uniform |
| Soldier       | Team     | Army   | Troop     | Logo             |
| Flora         | Jar      | Plant  | Pottery   | Potted Plant     |
| White Board   | Clothing | Coat   | Overcoat  | Suit             |
| Lighting      | License  | Couch  | Furniture | Plywood          |
| License Plate | Indoors  | LED    | Spotlight | Classroom        |

Show less

Fig. 5. Entities that are recognized in the latest video episode TNEU "Kaleidoscope of Events"

The next category of files that are usually synced from cloud storage systems is audio files. They can be divided into two categories: text from a voice recorder and phone conversations. In order to cover both categories, the software system uses the AWS Transcribe software, which provides an easy-to-use interface for interacting with audio files. At the entrance to the API, the file is sent in formats - mp3, ogg or wav, and the output will be text and timing.

An important advantage of using this software interface is the support of several speakers. An office worker will have the opportunity to select the text of the person whose record he wants to find. This functionality is provided by "uncontrolled" (unsupervised) machine learning. By entering several different voice records of different people, the algorithm of the neural network forms its guesses by breaking the audio tracks on the common features, thereby dividing the phrases that are similar to sound and assigning them to the people speaking them. Worth to note, that a few years ago, such technology was available only with a supporting video using separating audio using lips reading.

For other types of files that could not be divided into the media categories mentioned above, the Apache Tika software will be used to analyze even binary content (in total more than 1500 different MIME types of files) and extract media and information from them [4]. Examples of such information include dates, file creation, and people that created these files.

Apache Tika provides three ways to use: graphical user interface, command line software interface, and HTTP web server with a REST API.

The principle of building serverless systems involves freezing OS processes. But even here it's not an issue, as you can run Apache Tika server that receives the first request and then stops the process. The process state is saved and the new file process is restored, thereby eliminating the need to keep the running server running on the server permanently up to date. Even conventional bulky software from the 90s can be adapted to run in serverless environments to cut down the cost to the bare minimum.

Data transfer happens within the system in a private availability zone, no user information leaves the closed

network. Servers which run computations, do not have a network interface and even have no assigned IP address for access from the outside world. Latest OS patches are installed as soon as possible by cloud providers and require zero human intervention. As an example of security measures taken, let's remember a recent CPU vulnerability CVE-2017-5754 (Meltdown attack) which stressed out the whole world. Serverless environments were upgraded to a patched OS version within a more matter of hours after announcement without owners worrying patching a system manually.

In a knowledge management system, open source software plays a big role to identify keywords and key phrases. A group of programs for recognition and analyzing human speech comes from a range of products from the Stanford University, such as OpenNLP. It stands for Open Natural Language Processing. This framework provides an easy way to recognize keywords and entities from a text. Results are used in a search index to build filters and categories. For example, given an archive of library documents and analyzed contents, it is possible to share group thousands of materials into faceted search filters by the author, place of the event, time, country, and so on. A huge win for office workers when they search for a specific place or event in enormous blobs of text. And the second suggested use case of these rich text entities is building categories. Without training a model a system can group content related to a single date in history in one place. So it becomes easier to organize previously scattered content.

Consider an example of text analysis. The encyclopedia Wikipedia contains a fragment of the text from the article about the Taras Shevchenko National University of Kyiv for demonstration purposes.

The results of fragment recognition are shown in Fig. 6.

| Entity                                    | Category     | Count | Confidence |
|---|--------------|-------|------------|
| Ukraine                                   | Location     | 4     | 0.99       |
| Kyiv University                           | Organization | 4     | 0.99       |
| President of Ukraine                      | Person       | 3     | 0.85       |
| Taras Shevchenko National University      | Organization | 2     | 0.92       |
| 21 April 1994                             | Date         | 1     | 0.99+      |
| Leonid Kravchuk                           | Person       | 1     | 0.99+      |
| 25 November 1999                          | Date         | 1     | 0.99+      |
| Leonid Kuchma                             | Person       | 1     | 0.99       |
| Viktor Yushchenko                         | Person       | 1     | 0.99+      |
| 5 May 2008                                | Date         | 1     | 0.99+      |
| 29 July 2009                              | Date         | 1     | 0.99+      |
| University awards Junior Specialist       | Organization | 1     | 0.81       |
| 14 specialties                            | Quantity     | 1     | 0.99+      |
| More than 26 thousand students            | Quantity     | 1     | 0.99       |
| Approximately 1,645 postgraduate students | Quantity     | 1     | 0.93       |
| 125 PhD students                          | Quantity     | 1     | 0.99       |

Fig. 6. Key entities recognition

In a matter of minutes, gigabytes of information is transferred into a search index, where a user can filter out any document by several key criteria or by a person mentioned in the document.

### III. IMPLEMENTATION OF THE KNOWLEDGE MANAGEMENT SYSTEM

This knowledge management software is a web-based portal with a graphical user interface to provide an easy access to the search and the ability to drill down numerous search results through faceted filters, categories and suggestions.

A software implementation is based on JavaScript programming language and the Node.js software platform [5]. Many additional libraries, frameworks, and servers were included to parse and analyze files, such as Apache Tika, OpenNLP, Elasticsearch, Tesseract, GraphicsMagick, AWS Rekognition Video, and AWS Transcribe. Each file has a visual display for a quick and intuitive impression of what's inside a photo or a document.

### IV. CONCLUSION

The software system for synchronization of files from cloud storage and their analysis for the organization and contextual search on the basis of serverless computations is proposed and implemented. The developed software system is designed to handle the flow of files from cloud storage retrieved as a result of moving the directory structure tree hierarchy. The software system provides the ability to quickly search for tens of thousands of files. This solves a common problem among office workers who spend up to 500 hours a year searching for files. Such a large amount of information leads to information overload and fatigue from decision making.

The user feedbacks were analyzed on a sample of volunteers who used the system for 30 days. More than 90% of the results had positive feedback, so it is considered that the system has found its application and is meant to be useful.

The last but not the least, one important aspect of the developed system is its cost. Principles of maximum simplicity of software architecture design and low cost of maintenance were set in stone from the very beginning of development. As a result, the latest achievements in the field of computing were utilized, especially the principle of "serverless" computation. This approach hides a huge complexity from the knowledge system owner in terms of maintenance of own servers and data. Even database storage and search index implemented in serverless way ensure data persistence, availability, eventual consistency and idempotency by utilizing sharding and replication.

### REFERENCES

- [1] M. Stigler. *Beginning Serverless Computing*. Richmond, Virginia, USA: Apress, 2018, pp. 190-199.
- [2] A. Chaudhuri, Kr. Mandaviya, P. Badelia, S. K. Ghosh. *Optical Character Recognition Systems for Different Languages with Soft Computing*, 1st ed., Springer, 2016, pp. 2010-248.
- [3] C. Gurturk. *Building Serverless Architectures*. Istanbul: Packt Publishing, 2017, pp. 131-219.
- [4] Ch. Mattmann, Ju. Zitting. *Tika in Action*, 1st ed., Manning Publications, 2011, pp. 200-256.
- [5] A. R. Young, M. Harter. *Node.js in Practice*, 1st ed., Manning Publications, 2014, pp. 400-424.

# Software and Algorithmic Support for representation of CAD models in 2D von Neumann neighborhood

Yaroslav Sokolovskyy, Oleksiy Sinkevych

Ukrainian National Forestry University, Gen. Chuprynyk Str., 103, Lviv, 79057, UKRAINE,  
E-mail: sokolowskyyyar@yahoo.com, oleksiy1694@gmail.com

**Abstract:** In this paper we consider the system of initial theoretical positions, which are the basis for research on the search for variants of cellular automata in the study of thermo-mechanical processes in 3D objects of SolidWorks. These studies are carried out by representing 3D objects in the 2D von Neumann collations of the 1st order.

**Keywords:** CAD, lumber, cellular automata, von Neumann neighborhood 1st order.

## I. INTRODUCTION

The cellular automata is a discrete model, which is described by a set of a number of cells that forms a periodic lattice. This lattice allows you to determine the state of each of its cells at a specific time, and allows you to determine the state of its neighboring cells, the distance to which does not exceed the permissible limits.

Currently, cellular automata are often used in problems of modeling hydro and gas dynamic, evolutionary, behavioral, or different vibrational probabilistic processes, which in turn caused the comparative ease of implementation, which in turn gives better prospects for their future use.

The main direction of the research of cellular automata is the algorithmic solving of individual problems, which in one way or another use a two-dimensional coordinate system.

The result of this work will be the development of the concept of the use of cellular automata for the study of three-dimensional objects, as well as the creation of appropriate software, which will allow the use of this concept in conducting various types of experiments.

## II. FORMULATION OF THE PROBLEM

At the beginning of this work, we have a 3D model of fixed-sized piles of stakes that were assembled in the SolidWorks environment. This assembly of 3D model was created by using the developed and described in previous papers Software. [1], [2]

- This 3D model has the following initial parameters:
- Height, length and width of the stack;
- Number of lumber in the stack;
- The distance between lumber;
- Wood species;
- Dimensions of the pallet and gaskets;
- Temperature values (not necessarily the same);
- Moisture values (not necessarily the same).

In addition to the initial parameters of the loaded 3D model, it is necessary to specify the size of the drying space, which is filled with air mass. These sizes can be specified through the volume or through the values of its height, length and width. It

is also necessary to indicate the initial temperature and humidity of the drying space. You also need to specify a mathematical model that will allow you to change these values when changing time.

In carrying out this work is necessary to solve a number of problems, including:

- Calculate and determine the number of 3D cubes of the same size that can be diving through the model under investigation using the developed application software and the SolidWorks API;
- Convert each 3D cube to 2D neighborhood von Neumann 1st order and establish certain relationships between the faces of these 3D cubes using the above-mentioned dividing process;
- Visually present the established interconnections between the facets in the form of a corresponding scheme;
- Develop a set of rules that could use cellular automata to change the numerical values of temperature and humidity on the edges of developed 3D cubes;
- To develop classes with which we can implement the proposed layout of lumber of fixed size;
- It is necessary to develop a sequence diagram that will better understand the essence of the developed system classes.

Execution of all the above-mentioned tasks will allow all users to obtain the tools by which it will be possible to use the theory of cellular automata in determining the thermal and mechanical processes in the developed integrated CAD.

## III. DISTRIBUTION OF THE RESEARCHED OBJECT

The primary objective of this work is the allocation of a fixed size timber on 3D cubes, which then will be used for distribution in 2D neighborhood von Neumann.

For the distribution of lumber, it is necessary to take into account a number of rules, among which:

- It is necessary to divide the stack on a certain number of lumber, which will be divided into 3D cubes of the same size.
- The number of 3D cubes is limited and is determined by the cutoff density that is given in advance.
- The cut density should be sufficiently small to ensure that there are enough 3D cubes in the depth of the lumber.

In order to comply with the above rules it was decided to cut the lumber according to the scheme of uniform section with identical parts.

The selected lumber has the following dimensions: height and width of 100 mm, and length 1 m.

Further, according to the developed software application, the form of which is shown in Figure 1, the calculation of the

number and size of 3D cubes was performed. Consequently, with such dimensions of the lumber can be built 3D cubes, which can have 15 different sizes of external facet. If we take the largest dimension of the facet, which is 100 mm, then we will have only 10 cubes, which of course is not enough.

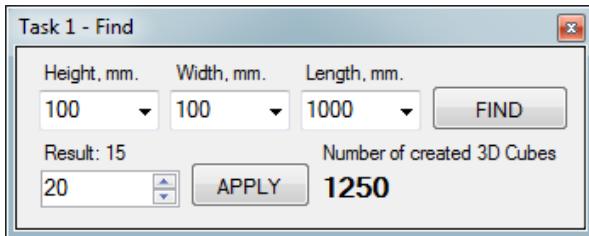


Fig. 1. View of the software application that is used to calculate the number of required 3D cubes

When selecting the size of facet at 0.25 mm, we get 640 million 3D cubes, the calculation of which, even on powerful computers take long. That is why it was decided to select 3D cubes, the outer edges of which have a size of 20 mm. With such dimensions of external facet, the total number of 3D cubes will be 1250 pieces.

After calculation of the required 3D cubes were made the timber section of a given size. This process is fully automated. This automation was achieved by using developed software application that uses capabilities of the SolidWorks API.

When we get 3D cubes, you need to assign them the value of the initial parameters of the stack. Assuming that the distance between the lumbers in the stack should be filled with airspace. Next, we need to divide the initial parameters into two categories, one of which will be assigned to the airspace, and the other formed 3D cubes.

The following parameters that will be assigned to 3D cubes include:

- Height, length and width are given according to the size of created 3D cubes, for example, 20 mm;
- Wood species;
- Initial temperature and humidity. If they are different, then it is necessary to uniform distribution along the length of a given timber.

The rest of the initial parameters are transmitted to the airspace surrounding the stack. Air space similarly will be presented as an array of 3D cubes of the same size. This representation will allow us to calculate a mathematical model that serves to change the parameters of the drying agent in space and time.

#### IV. CONVERSION OF 3D CUBES AND INSTALL DEPENDENCIES BETWEEN ITS FACETS

Consequently, when we having 3D cubes, it is necessary to make their transformation in order to represent them in the form of 2D squares. This step is very important and necessary, because with this transformation we will be able to use cellular automata

In general, one 3D cube has six facet, the numbering of which is shown in Figure 2.

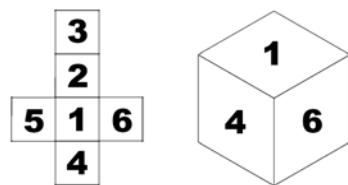


Fig. 2. Numbering of 3D cube facet

To represent cellular automata, we can use one of the two most popular 2D nodes, namely 2D Moore's area and 2D Von Neumann Fields. If we talk about the Moore countryside then it represents a set of eight cells on a square parquet, having a common vertex with this cell. In turn, the von Neumann occipital represents a set of four cells on a square parquet, having a common side with this cell. [3]

It is clear that in order to accomplish this task, the 2D field of von Neumann is best suited to the 1st order. According numbering of 3D cube facet and 2D neighborhood von Neumann in 1st order, we can get a graphical relationship facet of 3D cubes, which are shown in Figure 3.

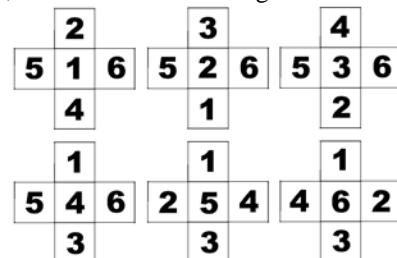


Fig. 3. View of 3D cube relationship facet by using 2D Von Neumann Fields in 1st order

In Figure 3, we can see how one 3D cube facet can affect another. Based on these relations was developed a general scheme for the internal and external facet of the created 3D cubes.

Visual representation of the principle of the scheme relationship between the external facets of 3D cubes, are shown in Figure 4.

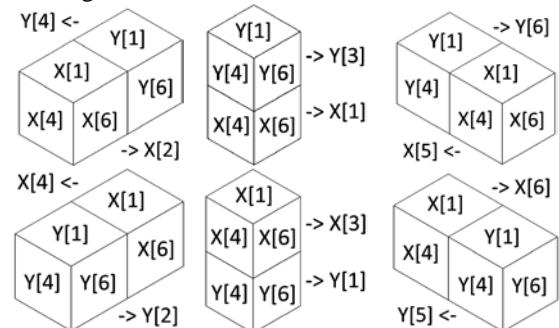


Fig. 4. Visual representation of the relationship scheme between the faces of different 3D cubes

The form of this schema is shown in Figure 5. The values of X [1-6] are the own facet of the selected 3D cube, and the value Y [1-6] is the facet of the outer 3D cube that is adjacent to the selected one. Consequently, each facet of the selected 3D cube can have a relation with only one facet of the outer 3D cube.

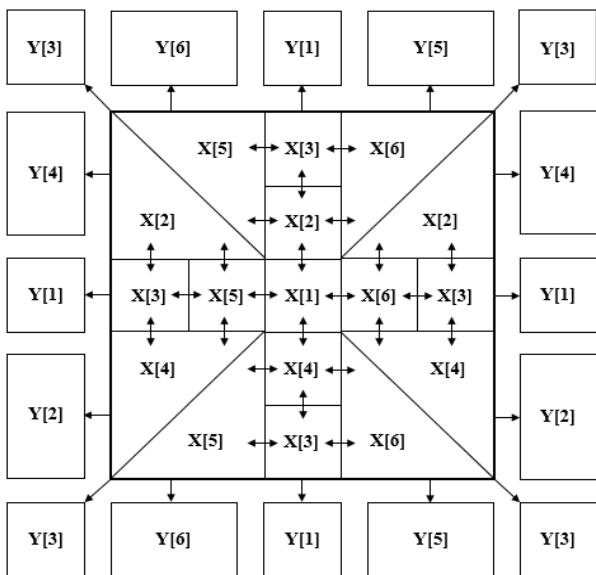


Fig. 5. View the general scheme for the internal and external facet of the created 3D cubes

## V. DEVELOPING THE RULES OF CONDUCT BASED ON CELLULAR AUTOMATES

Therefore, to accomplish the task by using cellular automata, it is necessary to develop some rules of behavior that should not violate the three basic components necessary for using of cellular automata, namely:

- The parallelism by which cell renewal occurs simultaneously and independently of one another;
- The locality in which the new cell value depends only on its previous value and the values of its cells around;
- Homogeneity, in which the cells must always be changed according to the same rules. [4]

To construct own rules for changing the values of the formed cells, we used: the relationship scheme, which is shown in Figure 4, the software application, the form of which is shown in Figure 1 and the three main components of the use of cellular automata.

In general, developed several rules, including:

- All 3D cubes must have faces of the same size, which in future will be prevent the presence of 2D cells of different sizes;
- All cells have their own initial values, which don't necessarily have the same numeric values (e.g. temperature or humidity);
- The faces of one 3D cube can only affect those faces of their own 3D cube, which are shown in Figure 3;
- Each facet of one 3D cube can directly affect only one facet of another 3D cube;
- Each 3D cube of the air environment can affect other faces of the same environment, or on the outer edges of those 3D cubes of lumber that are in direct contact with this environment;
- The value of the temperature and humidity of each face may vary according to the laws of thermodynamics, energy conservation laws, and the mathematical models developed and indicated in advance;

- The result of cellular automata can be mentioned desired temperature or humidity, or the passage of a specified time simulation.

## VI. DEVELOPMENT CLASS AND SEQUENCE DIAGRAM

To solve this problem, it is necessary to develop several basic classes. With these classes, you can implement a scheme of interconnections between the faces of 3D cubes.

The graphical relationship between the above-described classes is shown in Figure 6.

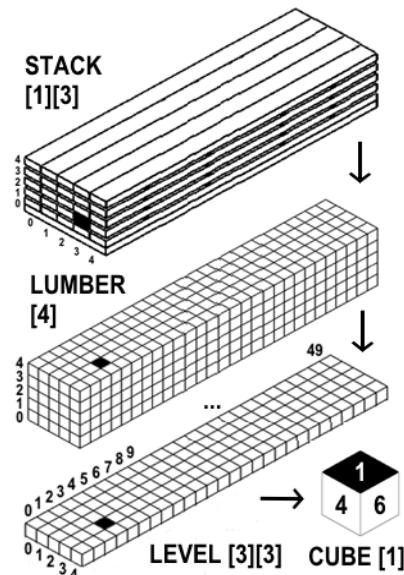


Fig. 6. Graphic interrelation between developed classes

The first class is called "Cube". This class contains information for each of the six faces of one cube. This information includes the numerical value of temperature, humidity and the general condition of each face. By itself, the state of the face allows us to know whether it is necessary to change its numerical value in the next iteration. In general, we can say that this class represents one of the smallest parts of the object being studied.

The second class is called "Level". This class is a two-dimensional array, which consists the elements of class "Cube". In essence, this class is an array of 3D cubes of the same level. If we recall the initial section of the object under study then it turns out this class has a dimension of 5x50 equal 250 cubes.

The third class is called "Lumber". This class is one solid timber. It is clear that this class is a one-dimensional array, which consists the elements of class "Level". Again, if we recall the initial section of the investigated object, then this class will have a dimension of 5, and this is 5 \* 250, which is equal to 1250 cubes.

The next class is called "Stack". It is clear that this class is a stack of lumber. This class is a two-dimensional array, which consists the elements of class "Lumber".

Class "3D\_Model" serves to download the original 3D model and its further use in the partition process.

The "Find" class is needed to find the lumber in the downloaded 3D model. This class gives us the ability to identify all possible section variants of the 3D model.

The next class is called "Cut" is required for the 3D model partition according to a given schema by using SolidWorks API. The last class "Calculation" serves for calculating numerical values on the edges of created 3D cubes. Before describing these classes, it was create the sequence diagram, which is shown in Figure 7.

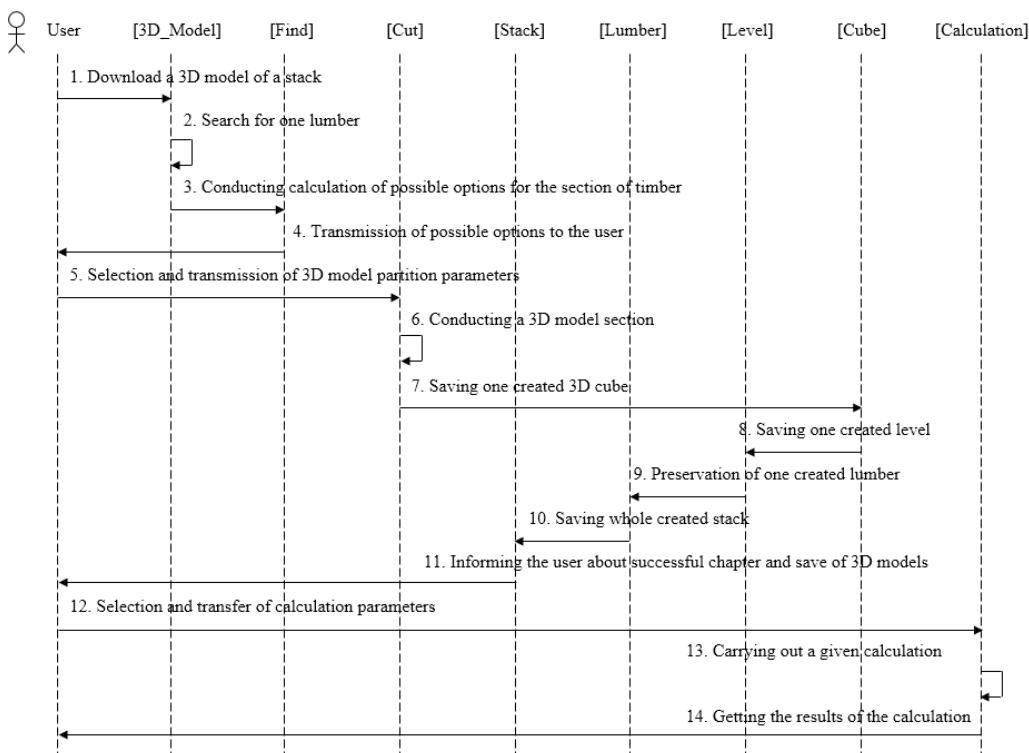


Fig. 7. View of the developed sequence diagram

### VIII. CONCLUSION

As a result of this work, was developed a concept for using the cellular automata for the study of three-dimensional objects.

In addition, was created a software with which you can download a 3D model of lumber piles and instantly calculate the number of possible variants of cutting each of the lumber into 3D cubes of the same size. The cutting process itself takes place using the SolidWorks API.

The resulting 3D cubes were presented in the form of 2D neighborhood of von Neumann 1st order. With this view, a relationship scheme was created for the inner and outer sides of the created 3D cubes. On the basis of established relationships was created, a certain set of rules for changing the numeric values that apply to the edges of 3D cubes.

Another important task that was accomplished in this work is the development of system classes, their further description, and a description of their interactions. As a visual representation of this was given a sequence diagram of the developed classes. In general, in this work, we examined the possibilities of using the theory of cellular

### VII. RESULTS

The result of this work is the creation the new software, which allows you to receive two-dimensional arrays of information that are necessary for using cellular automata. In addition, this software allows you to obtain arrays of information from the downloaded 3D models that are assembly in SolidWorks environment, by using early development application software. All of these processes were fully automated with the use of the SolidWorks Application Program Interface.

automata in studying models presented in a three-dimensional coordinate system. To study this concept, by conducting various types of experiments, was created an appropriate software in environment Microsoft Visual Studio C#.

### REFERENCES

- [1] Y. Sokolovskyy, O. Sinkevych, "Software for Automatic Calculation and Construction of Chamber Drying Wood and its Components", *Materials of the XII International Conference MEMSTECH 2016*, 20-24.04.2016, Lviv, Ukraine – Publisher Lviv Polytechnic, 2016. – pp. 209-213.
- [2] Y. Sokolovskyy, O. Sinkevych, "Calculation of the Drying Agent in Drying Chambers", *Materials of the XIV International Conference CADSM-2017*, 21-25.02.2017, Polyania, Ukraine – Publisher Lviv Polytechnic, 2017. – pp. 27-31.
- [3] Toffoli T., Margolus N. A Cellular Automata Machine. – M.: WORLD, 1991. – 280 p.
- [4] Wolfram S. A new kind of science. - Wolfram Media Inc., Champaign, Ill., USA.2002.

# Fast Contour Tracing Algorithm Based on a Backward Contour Tracing Method

Yuriy Batko, Vitalii Dyminskyi

Department of Computer Science, Ternopil National Economic University, UKRAINE, Ternopil, 8 Chehova str., email: [bum@tneu.edu.ua](mailto:bum@tneu.edu.ua)

**Abstract:** The article contains algorithm for detecting a connected objects contour. The algorithm was teste on digital biomedical images. The contours are use to separate objects from their backgrounds, to calculate the sizes of objects, to classify shapes and to find the feature points of objects).

**Keywords:** contour tracing, image analysis, contour chain, boundary following, pixel following.

## I. INTRODUCTION

Analysis form of objects plays a significant role in the course of many researches. In particular, changing the shape of the object can signal its transition from one state to another (for example, in the medical process of the course of the disease) [1]. In nature, the shape of an object is determine by its contours (cell walls, outside layer etc.).

Contour tracing is the stage of receiving a discrete signal describing the boundaries of the digital object.

The requirements for contour tracing algorithms are: reduction of storage space; reducing the time and complexity of further processing; obtaining informative features of the object.

The separate contour can take place in two ways: underlining the boundaries of an object by filtering the input image or by passing the inner contour of a homogeneous region.

The main algorithms for selecting the boundaries of the object are snake algorithm, Canny algorithm, filtration based on Sobel, Laplace, Prewitt and others. They are base on underscoring sharp drops of brightness, which are characteristic of the boundaries of objects. The result of their work is a set of unconnected areas. To obtain connected contour, it is necessary to carry out additional processing.

Algorithms for selection of areas: threshold segmentation, clustering, region-growing, watershed algorithm, block segmentation, etc. They are base on a union pixels in homogeneous regions based on a certain homogeneity criterion [2]. The result of their work is a set of homogeneous areas. To get the description of the contour of object, it is necessary to use contour tracing algorithms.

The following algorithms for passing the contour are know:

1) Square Tracing Algorithm [3] algorithm, the main advantage of which is simplicity. Contour tracing is base on two simple rules: if the value of the active pixel is equal to one (the active pixel is at the point belonging the object), then the left turn, otherwise, when the active pixel value is zero (the active pixel is at a point, which does not belong the

object), then right turn). The algorithm stops its work if it returns to the starting point.

2) Moore-Neighbor Tracing [4] - The algorithm is based on step-by-step verification of all adjacent points in order to find the next contour point. Stopping the algorithm occurs when it returns to the starting point.

3) "Radial Sweep" [5] - this algorithm is a modification of the previous one. Its main difference lies in the point of beginning of the bypass of the active pixel. In this algorithm, this is a point that was recognize as contour in the previous step of the algorithm, and not the point from which the transition to the active pixel occurred.

4) Theo Pavlidis's Algorithm [6] - The main idea of the algorithm is to use a group of three pixels to determine the next contour pixel. Verification is carry out by successive verification of adjacent points with a strictly defined sequence.

5) Snakes algorithm (active contour) [7] and Amoeba algorithm [8] are a group of algorithms base on finding contours by sequencing the image pixels to find the set of extreme (boundary) pixels. The algorithms stop their work if all possible pixels are searches or if there are no pixels that satisfy a certain condition.

6) Topological-hierarchical algorithms [9, 10] - a group of algorithms associated with tracking contours based on hierarchical relations between points. Algorithms of this group instead of markers use morphometric operations of finding points of overlap of several circuits for the purpose of their further separation.

These algorithms are used for analysis and describing the contour. Almost every programming language has libraries with implemented algorithms, for example, for Matlab, OpenCV etc.

Their main disadvantage is dependence on the complexity of the circuit and the stopping criterion. The algorithms are sensitive to microscopic objects, the contour of which contains a branch in one pixel thickness. This can lead both to the false end of the algorithms, and to the incorrect selection of the contour. A similar problem may occur if the micro-object consists of two or more parts that are connect only by single pixels. Another drawback of the algorithms is the imperfect criteria for stopping (returning to the starting point, passing a certain point several times), leading to incorrect work results.

## II. FAST CONTOUR TRACING ALGORITHM

A binary image with the selected objects is forming as result of the selection of objects in the images and the subsequent binarization procedure. These objects are

represent in the form of homogeneous areas. The geometric characteristics of these areas are important features for the analysis and perception of the image as a whole. Biological systems of visual perception, as shown by the research, use mainly the boundary of the border, rather than the separation of objects in brightness. In many cases, the most informative are the characteristics of the boundaries of the regions. To obtain the coordinates of the contour points, you must perform the path detection procedure by passing the boundary of the region.

We introduce the following definitions [11]:

Let  $Im$  be a binary digital image with  $N \times M$  pixels, where the coordinate of the top-leftmost pixel is  $(0,0)$  and that of the bottom-rightmost pixel is  $(N-1, M-1)$ . If  $Im(x, y) = 1$ , then this point belongs to the object, otherwise - the background.

Start pixel  $Im_s(x, y)$  is the pixel from which the object boundary begins to go. The choice of the starting point is arbitrary, for example, the extreme left-top pixel belonging to the object. The final pixel  $Im_e(x, y)$  is the pixel, upon which the algorithm ends its operation. Active pixel  $Im_a(x, y)$  - pixel located in the middle of the marking grid. Neighboring pixel  $Im_n(x, y)$  is a pixel that borders on the active pixel.

Contour Pixel  $Im_k(x, y)$  is a pixel belonging to the object's contour. Background pixel  $Im_f(x, y)$  is a pixel that belongs to the background image. The neighboring contour  $Im_{kn}(x, y)$  pixel is a pixel that belongs to the contour of the object and borders on the active pixel.

After analyzing the advantages and disadvantages of existing methods, the following algorithm for contour tracing has been developed with the ability to cut off the informational branches of the object's path, which is represented by a sequence of steps:

Start pixel is searching  $Im_s(x, y)$ .

Finding a neighboring contour pixel  $Im_{kn}(x, y)$  clockwise and a neighboring contour pixel  $Im_{kn}(x', y')$  counterclockwise.

If the adjacent contour pixels match  $Im_{kn}(x, y) = Im_{kn}(x', y')$ , then the active pixel is recognized as background  $Im_f(x, y)$  and excluded from further processing and the transition to p. 8 is carried out.

If the resulting contour pixels do not match  $Im_{kn}(x, y) \neq Im_{kn}(x', y')$ , then the start pixel is also recognized as the endpoint  $Im_s(x, y) = Im_e(x, y)$ . The neighboring contour pixel obtained in step 2 is entered into an array of contour pixels and assigned to it an active pixel label  $Im_{kn}(x, y) = Im_a(x, y)$ .

The next neighboring contour pixel is searching  $Im_{kn}(x, y)$ . The sequence of checks of adjacent contour points is based on a marking grid. Pixels are checked clockwise. The starting checking position  $d$  is determined by  $(d'+2) \bmod 8$ , where  $d'$  is the position from which the active pixel was found  $Im_a(x, y)$ . Contour tracing is illustrated on the Fig. 1. The search ends when you find the next contour pixel or to check all neighboring contour pixels.

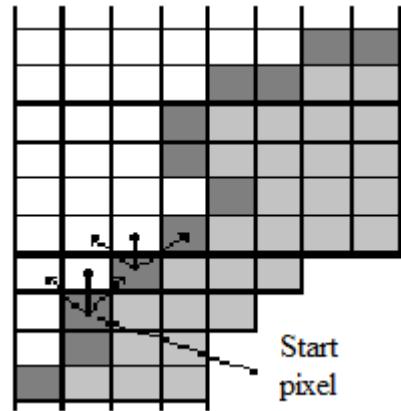


Fig.1. Contour tracing for steps

If the neighboring pixel is a contour and does not match the end pixel, then it is entered into an array of contour pixels  $Im_{kn}(x, y) = Im_a(x, y)$  and assigned an active pixel tag.

If an adjacent pixel is found in the previous search steps, but does not match the end pixel  $Im_{kn}(x, y) \neq Im_e(x, y)$ , then the active pixel is considered to be non-informative, removed from the array of outline points, indexed as the pixel of the background. The active pixel status is assigned to the previous contour pixel.

If the resulting contour pixel coincides with the end pixel  $Im_{kn}(x, y) = Im_e(x, y)$  and the number of points in the contour is greater than one, then complete.

The time of algorithm work can be reduced by the parallel tracing both clockwise and opposite. In step 3, a parallelism of verification was performed. The starting point for validation is calculated as  $(d''-2) \bmod 8$ , where  $d''$  is the position from which the counterclockwise active pixel was found  $Im_{ac}(x, y)$ .

When  $Im_a(x, y)$  or  $Im_{ac}(x, y)$  is found already marked contour pixel, then a label type check is performed. If the contour label was, install by them in the previous step, and then such a pixel is considered to be non-informative. If the contour mark was, install by the other side, then the contour is considered to be found.

To obtain a connected path, the points obtained by moving against the clock are successively added to the set of contour points obtained by direct motion.

In order to evaluate the operation of the contour tracing algorithm, classify the complexity of the contours of objects Table I. Simple contours are contours that can be described

(approximated) using 3 to 10 straight lines. Normal contours are contours that are approximated by more than 10 straight lines. Complex contours are described by a array of more than 10 straight lines, between which there are both sharp and dull angles. The approximating maximum error value is 2 pixels.

TABLE I. OBJECT CONTOUR CLASSIFICATION

| Simple contours        | Normal contours             | Complex contours  |
|------------------------|-----------------------------|---|
|                        |                             |   |
| 3 to 10 straight lines | more than 10 straight lines | more than 10 straight lines and different angles between them |

The results of contour tracing algorithms for normal contour are showing in Table II.

TABLE II. COMPARISON CONTOUR TRACING ALGORITHMS FOR DIFFERENT PARAMETERS

| Algorithms                              | 4- Connectivity contour | 8- Connectivity contour | Number of inspections |
|---|-------------------------|-------------------------|-----------------------|
| Square Tracing Algorithm                | +                       | -                       | 1-4                   |
| Moore-Neighbor Tracing                  | +                       | +                       | 1-7                   |
| Redial Sweep                            | +                       | +                       | 1-8                   |
| Theo Pavlidis's Algorithm               | +                       | +                       | (1-3)x3               |
| Fast Backward Contour Tracing Algorithm | +                       | +                       | 1-6                   |

The advantages of this algorithm are:

- work with 8-connected contour;
- Independence from the choice of the start pixel;
- high speed by reducing points for analysis, for example, in comparison with the "Redial Sweep" algorithm can make up to 25% (Table II);

- The process of the algorithm's operation can be simply parsed into several streams, which reduces the processing time of the image. The accuracy of the algorithm results does not decrease in parallel processing;

- the possibility of a rollback of work. It should be noted that the principle of returning the operation of the passage algorithm has not been used previously. Using the rollback of work prevents the looping of the algorithm, the rejection of informational pixels and the possibility of correct work with complex contours.

Testing algorithm for correct work will be performed on digital images with the cytological objects. Cytological objects have different complexity of contour line, therefore they are optimally suited for testing. An example of testing is shown on Fig.2.

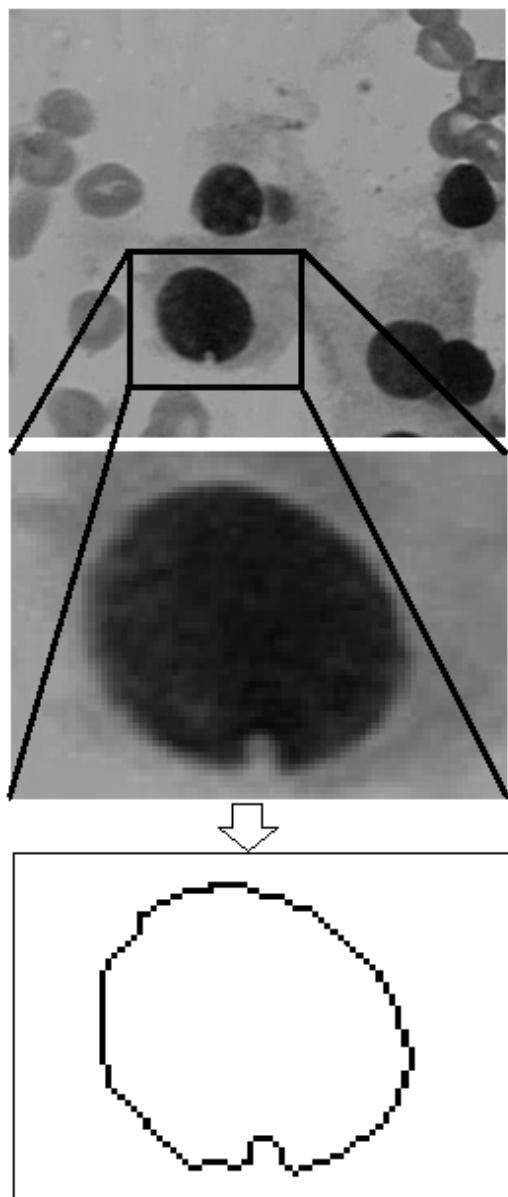


Fig.2. Contour tracing with Fast Backward Contour Tracing Algorithm

It is possible to form a connected contour of object, without rejecting the informative pixels. In this case,

informational pixels do not belong to the background of the image, but remain in the matrix of the outline points with the appropriate label assignment to them.

Results of simulation of contour tracing algorithms on the images chess figures, which including object with normal contour are showing in Tables III and IV.

TABLE III. CONTOUR TRACING WITH FAST BACKWARD CONTOUR TRACING ALGORITHM

| Image object | Contour object |
|--------------|----------------|
|              |                |
|              |                |
|              |                |
|              |                |

TABLE IV. COMPARISON CONTOUR TRACING ALGORITHMS

| Algorithms                              | Perimeter, pixel | Perimeter, % | Number of inspections | Number of inspections, % |
|---|------------------|--------------|-----------------------|--------------------------|
| Square Tracing Algorithm                | 683              | 100%         | 976                   | 100%                     |
| Moore-Neighbor Tracing                  | 523              | 77%          | 1356                  | 139%                     |
| Redial Sweep                            | 523              | 77%          | 1473                  | 151%                     |
| Theo Pavlidis's Algorithm               | 523              | 77%          | 1289                  | 132%                     |
| Fast Backward Contour Tracing Algorithm | 447              | 65%          | 1089                  | 112%                     |

The Square Tracing Algorithm was chose as standard because it is easy to implement and the result of its work is always productive.

### III. CONCLUSION

The contour tracing algorithm is allows to discard the less informative branches for further analysis, since it provides for a rollback process. However, if it is necessary to obtain a complete connected contour of object, the possibility of processing small informative branches is foreseen.

Obtained connected object contour in the form of a sequence of points coordinates can be further analyzed. For example, to reduce the amount of memory for storing the contour can perform an approximation procedure.

### REFERENCES

- [1] A. Kazlouski; R. K. Sadykhov, "Plain objects detection in image based on a contour tracing algorithm in a binary image", *Innovations in Intelligent Systems and Applications (INISTA) Proceedings, 2014 IEEE International Symposium, 23-25 June 2014*, DOI: 10.1109/INISTA.2014.6873624.
- [2] O. Berezsky, G. Melnyk, Yu. Batko, O. Pitsun, "Regions Matching Algorithms Analysis to Quantify the Image Segmentation Results", *Proceedings of the 11th International Scientific and Technical Conference CSIT, 2016*, pp. 200-203.
- [3] William K. Pratt, Digital Image Processing: *PIKS Inside, Third Edition*, John Wiley and Sons, Inc., New York. – 2001–736p.
- [4] V. Hlavac, M. Sonka, R. Boyle, Image Processing, Analysis, and Machine Vision, International Student Edition. *Thomson Learning, part of the Thomson Corporation* – 2008 – 829p.
- [5] P.Rajashekhar, "Evaluation of Stopping Criterion in Contour Tracing Algorithms" (*IJCSIT International Journal of Computer Science and Information Technologies*. – Vol. 3 (3). – 2012. – P. 3888-3894.
- [6] T. Pavlidis, Algorithms for Graphics and Image Processing, *US, Rockville, Maryland: Computer Science Press*, 1982. – 438 p.
- [7] M. Kass, A. Witkin, D. Terzopoulos, Snakes: active contour models, *Int J Comput Vis* – 1988- pp. 321–331.
- [8] G. Iannizzotto, L. Vita, Fast and accurate edge-based segmentation with no contour smoothing in 2-D real images, *IEEE Trans Image Process* 9- 2000-1232–1237.
- [9] P. Kuagoolkijgarn, P. Koomsap, N. Chansri, A new algorithm for tracing nests of interconnected contours, *The International Journal of Advanced Manufacturing Technology* September 2010, Volume 50, Issue 5–8, pp 717–727.
- [10] P. Koomsap, N. Chansri, Topological hierarchy-contour tracing algorithm for nests of interconnected contours, *Int J Adv Manuf Technol* (2014) 70:1247–1266.
- [11] Yu. Batko, O. Berezsky "Algorithm of determination of image contours of biological nature", *Proceedings of international conference "Modern problems of radio engineering, telecommunications and computer science" TCSET 2006, 28 February – 4 March 2006, Lviv, Ukraine.* – 2006. – pp. 642–644.

# Computer and Network Security

# Cloud Infrastructures Protection Technique Based on Virtual Machines Live Migration

Jan Fesl<sup>1</sup>, Vineet Gokhale<sup>1</sup>, Marie Doležalová<sup>1</sup>, Jiří Cehák<sup>1</sup>, Jan Janeček<sup>2</sup>

1. Institute of Applied Informatics, Faculty of Science, University of South Bohemia, CZECH REPUBLIC, České Budějovice, Branišovská 31a, email: [jifesl.vgokhale.dolezm01.jcehak@prf.jcu.cz](mailto:jifesl.vgokhale.dolezm01.jcehak@prf.jcu.cz)

2. Department of Computer Systems, Faculty of Information Technology, Czech Technical University in Prague, CZECH REPUBLIC, Prague, Thákurova 6, email: [janecek@fit.cvut.cz](mailto:janecek@fit.cvut.cz)

**Abstract:** Cloud computing and virtualization are very popular research areas of the last years. The rising count of running virtual machines brings new opportunities for the malware or intrusion tools. Nowadays, there exist various ways how to detect the potential attack activities. Our paper evaluates such techniques and provides an efficient solution how to link the anomaly detectors and hypervisors. This solution provides the way how to eliminate the problems without having to shutdown the running virtual machines.

**Keywords:** IDS, IPS, hypervisor, cloud, security, live migration, virtual machine.

## I. INTRODUCTION

The distributed virtualization infrastructures (DVI) are the engines of powerfull data centres, which create the executive parts of cloud environments. The typical virtualization architecture, which is depicted in figure 1, consists of the management node (MN), virtualization nodes (VN), data storage (DS) and interconnection networks (IN). Virtual machines (VM) run on the virtualization nodes, which contain the hypervisors. Hypervisor is the abstract layer between the physical and virtual hardware. There are more principles of virtualization techniques and their detail description can be found in [1]. Nowadays, the most power efficient hypervisors are based on direct cooperation with the hardware using dedicated instruction set.

$$DVI = \{MN, VN, DS, IN\} \quad (1)$$

The virtual machines can be moved during the runtime between all infrastructure nodes. The motion of the virtual machines allows the live migration technique. The detail info about it, can the reader find in [2]. The virtual machines have some limitations like allowed count of CPUs, size of operating memory, size of virtual hard drive or maximal network throughput. All parameters are managed by the hypervisor. The virtual machine lives in its own world and shares its resources with other virtual machines. The level, on which the physical and virtual computers are the same, is the network level, because both types communicate with outer world via the same packets. The virtual computers contain the same operation system like the physical, therefore they can suffer from the same security issues. The virtual machines are interconnected at the same virtualization nodes by virtual networks. The virtual networks can bridge some virtual machines with others or fully separate the network

traffic between them.

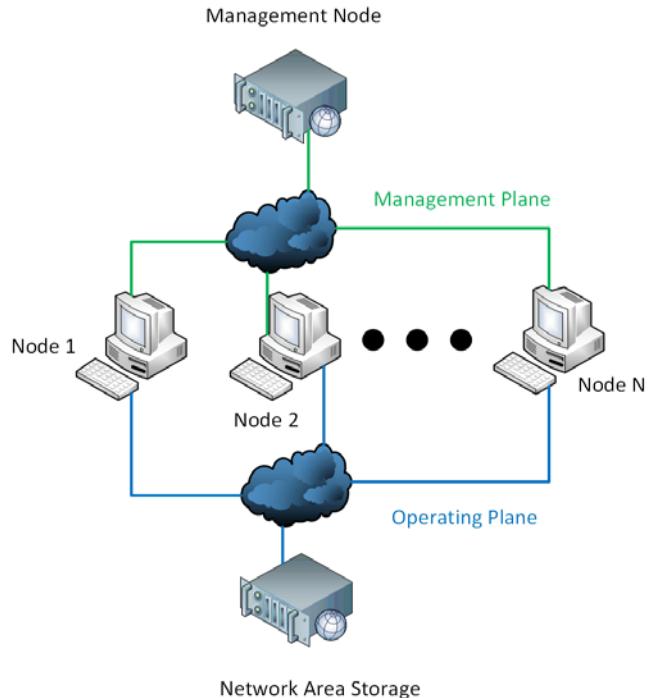


Fig. 1. Typical environment of the distributed virtualization infrastructure.

## II. SYSTEMS FOR INTRUSION DETECTION AND PREVENTION

The infected virtual machines can be the sources of potential network attacks. From the network OSI model point of view, the attack types can be divided into L2 and L3 groups. The first group is targeted on the computers and devices placed in local segments, the second group can affect all local and remote devices, which are connected to the computer network.

In the data center environment, the greatest problem for the administrator is, that it is not possible to permanently apply strict firewall rules like limit TELNET, SSH, RDP etc. services, because the users require to install and use various network services. The data centers can't also require from an user a paid antivirus/antimalware protection or firewall. The only one acceptable solution is to suppress the potential issue, but not to drop all other services. The traditional firewalls are

highly efficient, but strictly static and are not very good suitable for the data centers. The efficient way is to use more intelligent intrusion detection (IDS) and intrusion prevention (IPS) systems. Intrusion detection systems (IDS) are an essential component of protecting computer systems and network. To detect computer attacks and provide the proper response this is the main aim of IDS. An IDS is defined as the technique that is used to detect and respond to intrusion

activities from malicious host or network. There are mainly two categories of DSs, network based and host based. IDS is key to detect and possibly prevent malicious activities. In the case, that the issue has been discovered, the administrator is informed by sending the message. Such systems have had the long tradition, but were primary proposed for the common computer networks. The basic structure of such system is depicted in figure 2.

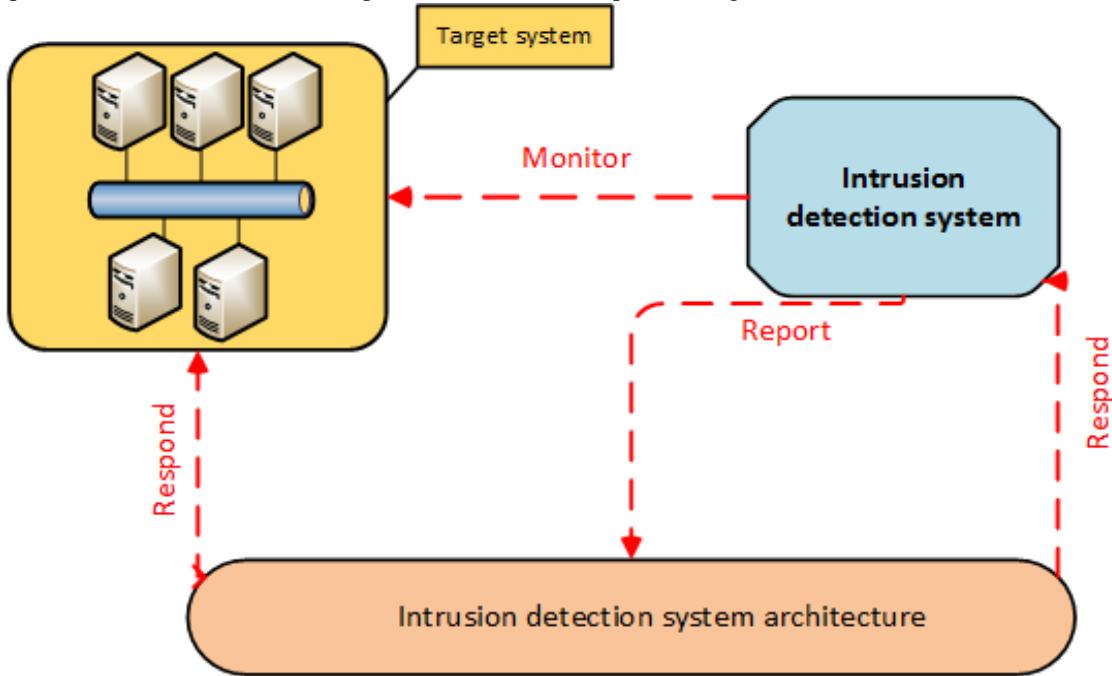


Fig. 2. Main components of the IDS system, which could be used as data center protection system.

The really comprehensive overview of this problematics including algorithms description is listed in [2] and some other in [3]. The basic detection techniques used in IDSs are as follows.

#### *Signature matching*

This techniques identify attacks by matching network packet contents against specific attack signatures. These signatures are created using already identified and well-described attack samples, which is time consuming and can take from couple of hours up to several days, which gives attackers plenty of time for their criminal activities. The biggest weakness of this solution is that it is detecting only known attacks, which can be due to smart evasion techniques used by malware limiting. With the growing proportion of encrypted traffic, use of self-modifying malware, and other evasion techniques, the use of a detection technique tailored to catch predefined set of attacks is becoming irrelevant.

#### *Anomaly-based detection*

This concept tries to decrease the human work (e.g. manual creation of signatures) by building a statistical model of a normal behavior and detect all deviations from it. This enables to detect new, previously unknown attacks provided that their statistical behavior is different from that of the normal traffic. While anomaly-based methods are attractive conceptually, they have not been widely adopted. This is

because they typically suffer from comparatively higher false alarm rate (not every anomaly is related to the attack) rendering them useless in practice, since network operator can analyze only few incidents per day [3]. That is why the signature based IDS are still widely used even they are not able to detect new types of attack nor to find anomalous behavior of the network users.

In the literature [4], there are known five types of IDS systems, which can be used in the cloud computing.

#### *Host based intrusion detection system (HIDS)*

Such system is an intrusion detection system that monitors and analyzes the information collected from a specific host machine. HIDS running on a virtualization node detects intrusion for the virtual machine by collecting information. HIDS observes modification in host kernel, host file system and behavior of the program. Upon detection of deviation from expected behavior, it reports the existence of attack. The efficiency of HIDS depends on chosen system characteristics to monitor. Each HIDS detects intrusion for the machines in which it is placed. With respect to Cloud computing, HIDS can be placed on a host machine, VM or hypervisor to detect intrusive behavior through monitoring and analyzing log file, security access control policies and user login information.

#### *Network based Intrusion Detection System (NIDS)*

It is an intrusion detection system that tries to detect malicious activity such as DoS attacks, port scans or even

attempts to crack into computers by the network traffic monitoring. The information collected from network is compared with known attacks for intrusion detection. NIDS has stronger detection mechanism to detect network intruders by comparing current behavior with already observed behavior in real time. NIDS mostly monitors IP and transport layer headers of individual packet and detects intrusion activity. NIDS uses signature based and anomaly based intrusion detection techniques. NIDS has very limited visibility inside the host machines. If the network traffic is encrypted, there is really no effective way for the NIDS to decrypt the traffic for analysis. NIDS can be deployed on edge node interacting with external network. However, it has several limitations. It can't help when it comes to attack within a virtual network that runs entirely inside the hypervisor. In cloud environment, NIDS must be installed by the service provider.

#### *Distributed Intrusion Detection System (DIDS)*

This system consists of many IDS over a large network, which communicate together or with a central server that

enables network monitoring. The intrusion detection components collect the system information and convert it into a standardized form to be passed to the central analyzer. Central analyzer is a machine that aggregates information from multiple IDS and analyzes the same. Combination of anomaly and signature based detection approaches are used for the analysis purpose. DIDS can be used for detecting known and unknown attacks since it takes advantages of both the NIDS and HIDS, which are complement of each other. In cloud environment, DIDS can be placed on the virtualization nodes.

#### *Hypervisor based intrusion detection system (HIDS)*

The proposal of such system is still more in theoretical surface then in some real implementation. From the real world, Hyper-V hypervisor [5] contains the L3 firewall, which can be used for blocking of some attacks. The principle is depicted in figure 3. The greatest trouble is its implementation, which can cause the rapid virtualization technology performance decreasing.

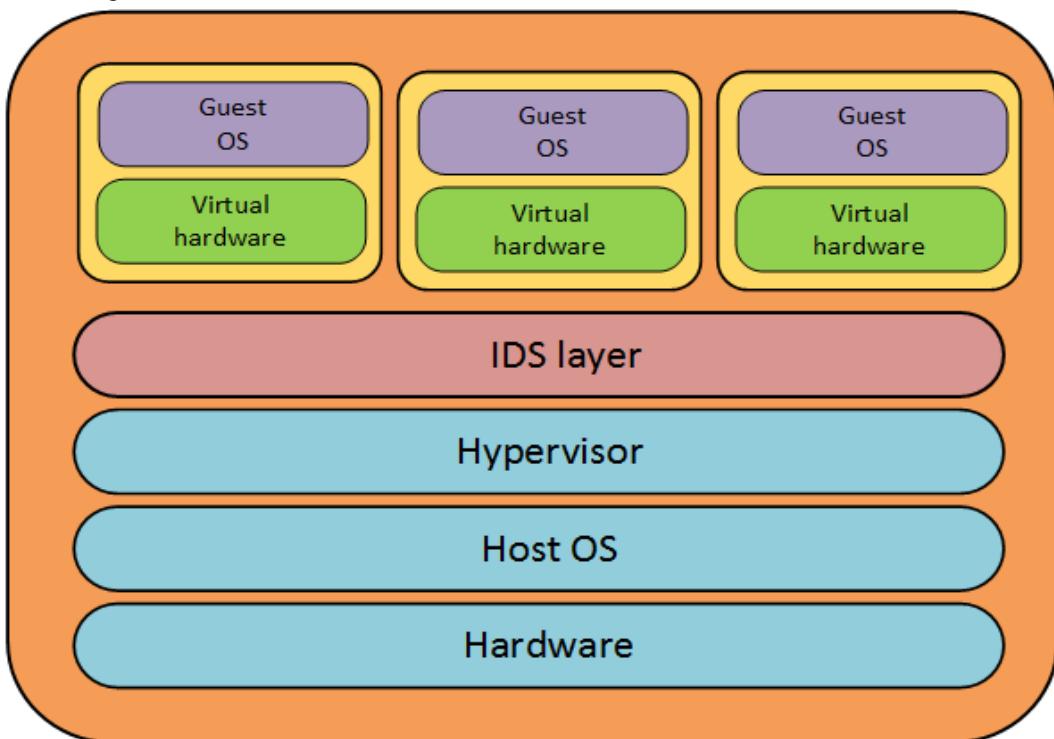


Fig. 3. Hypervisor based IDS, integration with the operating system.

#### *Virtual machine introspection (VMI) based IDS*

VMI is the main idea behind out-of-box intrusion detection. VMI is a technique of inspecting VM state by moving the inspection module outside of the VM. The software running inside the guest system is analyzed externally to detect any intrusion. One advantage of this technique is that malware detection continues to work unaffectedly even in the presence of an intrusion. This capability is missing in HIDS and NIDS. In the case of a compromise, HIDS starts reporting falsely while NIDS has limited visibility. More info can be found in [6] or [7] and [8].

### III. PROTECTION TECHNIQUE BASED ON VIRTUAL MACHINES LIVE MIGRATION

The basic idea behind this technique is as follows. The virtualization infrastructure with an active IDS can be divided into three smaller elastic independent groups. The groups differs by the level of network traffic filtering – {A,B,C}. The group A contains virtualization nodes with VMs with no traffic filtering. The group B contains virtualization nodes, which are L3 protected for some outgoing L3 attacks e.g. SSH, TELNET or SMTP. The type of the suppressed traffic for concrete VM is strictly evaluated by the IPS. The group C is

called “Quarantine” and has at L2 strictly separated and filtrated traffic off all virtual machines. The specific outgoing L3 traffic is filtered as well - it can be the most of the L3 outgoing traffic, depends on IPS decision. The users, who manage VMs in the group C, can connect to them via the specific terminal service. The IDS directly cooperates with the hypervisors of all virtualization nodes. If the IDS detects some issue, it gives a command to the specific hypervisor which migrates the affected VM to the virtualization node, which is the member of other more secured group. Before the migration of VM, all restrictions proposed by IPS are activated on the destination node. Live migration serves for isolation of potential malicious and health virtual machines. The principle is depicted in figure 4. This principle is further summarized in the following 5 step algorithm.

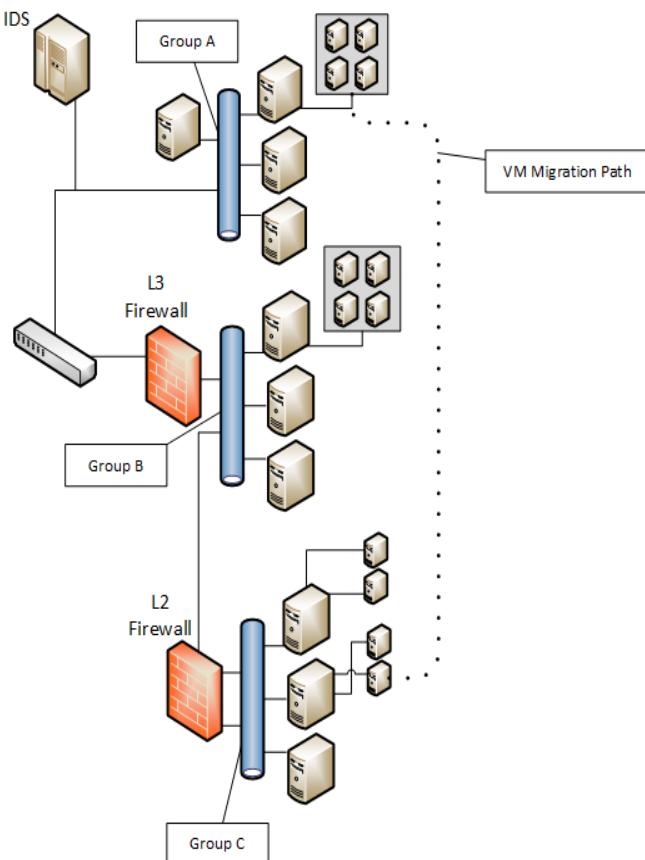


Fig. 4. Live migration protection technique principle.

#### The 5-step protection algorithm (managed by IDS)

$V$  is the set of new potential affected virtual machines.  $R$  is the set of recommendations, which means the new running groups for  $V$ -contained VMs, e.g. VM runs in A, but has a security issue, than IDS recommends move this VM from A to B,  $R(\{VM, A \rightarrow B\})$ .

1. Detect all possible problematic running VMs and store them in  $V$  and their recommendations in  $R$ .

2. For all VMs in  $V$  find the destination nodes (DN) according to  $R$  values for their migrations.
3. Apply all L2/L3 protection rules on DN for all VMs in  $V$ .
4. Give a command for live migration of all VMs in  $V$  to all involved hypervisors.
5. Give a message to all administrators of the affected VMs which VMs were migrated.

The algorithm can be more generalized for the positive reevaluation which means, that if some issue on the VM disappears, the VM can be back migrated to less strict group.

#### IV. CONCLUSION

We proposed the new technique which is able to help by the protection of the distributed virtualization network infrastructures. Its deploying depends on the election of the proper IPS element. The greatest benefit of such solution stands in the runtime protection without supressing more services, than it's necessary. The solution requires for the deployment such hypervisor type, which is able to migrate the VMs during the runtime. The advertised technique is suitable for all L2 or L3 network attack types.

#### REFERENCES

- [1] J. Fesl, "Virtual distributed systems and their application", dissertation thesis, Czech Technical University in Prague, 2017, pp. 10-27
- [2] Fesl, J., et al., "Live Migration of Virtual Distributed Computing Systems", *International Journal of Innovative Computing Information and Control. (IJICIC)* 2015, Vol. 11, Issue 3
- [3] J. Grill, "Combining network anomaly detectors", dissertation thesis, Czech Technical University in Prague, 2016, pp. 3-17
- [4] F. Alruwaili, T. Gulliver, "CCIPS: A Cooperative Intrusion Detection and Prevention Framework for Cloud Services", *International Journal Latest Trends of Computing*, Vol. 4, Issue 4, 2014
- [5] A. Lownds et al., "Windows Server 2012 Hyper-V Installation and Configuration Guide"
- [6] A. Riaz et al., "Intrusion Detection Systems in Cloud Computing: A Contemporary Review of Techniques and Solutions", *Journal of Information Science and Engineering*, Issue 4, 2016
- [7] T. Hwang, Y. Shin, K. Son, H. Park, "Design of a Hypervisor-based Rootkit Detection Method for Virtualized Systems in Cloud Computing Environments", *AASRI Winter International Conference on Engineering and Technology*, 2013
- [8] Y. Tayyebi, D. Bhilare, "Cloud security through Intrusion Detection System (IDS): Review of Existing Solutions", *International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)*, Vol. 4., Issue 6, 2015

# Omnifactor Authentication

Libor Dostálek<sup>1</sup>, Iva Dostálková<sup>2</sup>

1. Department of Applied Informatics, Faculty of Science, University of South Bohemia, CZECH REPUBLIC, Ceske Budejovice, Branisovska 1760, email: [dostalek@prf.jcu.cz](mailto:dostalek@prf.jcu.cz)

2. Department of Mathematics and Biomathematics, Faculty of Science, University of South Bohemia, CZECH REPUBLIC, Ceske Budejovice, Branisovska 1760, email: [dost@prf.jcu.cz](mailto:dost@prf.jcu.cz)

**Abstract:** At present, many authors looking for new and new authentication methods [1] [2], which their authors consider that, are stronger and stronger. Individual one-factor authentication methods combine to other, and thus multi-factor authentication will arise. Today's applications typically allow users to choose from a set of supported authentication methods. Users can choose to use one or a combination of multiple authentication methods. We say that users have available omnifactor authentication. The question is how to quantify the strength of the selected methods in kind of omnifactor authentication. This is important to determine if is the authentication sufficient for the required service (required information). This article offers the answer to this question by quantifying authentication based on knowledge, ownership and inherence factors.

**Keywords:** authentication, multifactor authentication, omnifactor authentication.

## I. INTRODUCTION

Authentication is the entity's identity verification process. This process is carried out by a verifier who guarantees that the entity has a declared identity (Figure 1). The quality of this warranty depends on the particular authentication process.

We distinguish entity authentication and message authentication. The difference is in the time perspective. Authentication of the message (e.g. by electronic signature) gives no guarantee as to when the message was created. Instead, entity authentication includes proof of identity of the applicant as a rule through current communication with the verifier.

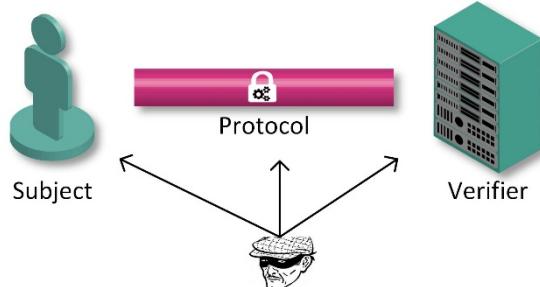


Fig. 1. Fundamental roles in authentication process.

An example of an authentication process is the process by which a user logs on to the application using a username and password.

The secondary effect of the authentication process may be

the fact that during the authentication of the entity, cryptographic material will also be generated to serve for subsequent communication.

The ways in which someone may be authenticated fall into three categories, based on what are known as the factors of authentication:

1. The subject something knows - knowledge factors – e.g. password, private or secret key, shared secret, etc.
2. The subject has something - ownership factors – e.g. smart card, one-time passwords token, etc.
3. Subject something is or does - inherence factors – e.g. fingerprint dynamic biometric signature, digital footprint, etc.

Multi factor authentication grant access only after successfully presenting two or more factors (Figure 2).

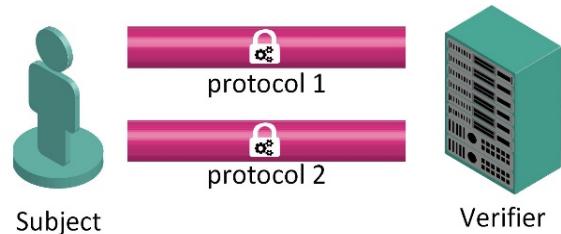


Fig. 2. Multi factor authentication.

It is important that different authentication factors are used. E.g. using two passwords does not improve the authentication quality. Authentication factors may vary:

- Different cryptographic material.
- Different authentication scheme.
- Different communication protocol.
- Different communication channel.
- Different verifier.

It is also important that the authentication factors are intertwined. If they are not, then the attacker makes work easier because the attacker can first deal with breaking one authentication factor and then another. However, this cannot always be achieved. E.g. if an entity is already authenticated (for example, it has brought authentication from Facebook) and it turns out that stronger authentication (e.g. smart card) is required for the operation, then it is usually re-authenticated only with a stronger chip scheme independent of the original authentication. The basic disadvantage of biometric person's characteristics is that they cannot be revoked and subsequently altered in the case of abuse. E.g. if an attacker obtains a dynamic biometric signature from the subject, then the subject

can never use a dynamic biometric signature without compromising its misuse.<sup>1</sup>

## II. RELATE WORKS

The inheritance authentication factors are usually considered as based on the biometric properties of the subject, i.e. verification of the person's identity based on measurable physiological or behavioral characteristics, unique and relatively unchanging for the subject.

Authentication takes place based on the input pattern match pre-stored template e.g. in the database. Matching cannot be absolute - probably, it may be an attack. Authentication is confirmed if the matching exceeds a predetermined threshold.

So far, we have considered authentication as it is used, for example, when logging in to FTP or Telnet server. However, at present, logging is part of wider communication, for example, when a user logs in to a web server. The communication for displaying the logon page transmits a large amount of data and the logged-on user leaves the digital footprint. A digital footprint can be used for authentication itself or can serve as another authentication factor. Interestingly, the information extracted from the digital footprint can not only amplify but also weaken the resulting authentication. Weaken, if are detected potential attack.

If we want to use the digital footprint for authentication, then:

1. From a digital track, we must be able to identify users. The easiest way to do this is to save the user's identification to cookies.
2. Upon subsequent authentication, we can determine the degree of match information in the current digital footprint with information in the previous digital footprints of the same user.
3. If we find a mismatch of the current digital footprint with the previous digital tracks, then we can ask the user for additional (secondary) authentication.

The Digital Footprint has similar features to biometric characteristics. The key is to be able to identify user from the digital track. However, even when digital footprint are able to identify users with a certain probability, it can be useful in practice. It can be used, for example, to distribute a customized ad.

Risk based authentication is called authentication based on the calculation of the likelihood of attack against this authentication. The patent [1] deals with user authentication in the kind of an Internet service provider's application. This can be, for example, an e-shop, electronic banking or eGovernement application. The client during authentication and subsequent downstream communication leaves a lot of information in the communication channel - leaving a digital footprint. The principle is as follows: the communication flow between the subject and the verifier is duplicated. The duplicate of the communication stream is redirected to the Risk Engine, which then evaluates this information as the input information for calculating the risk score (Figure 3). The risk score can then serve as input:

- For risk-based authentication that is used in this paper.
- Fraud detection system, which is used to detect cyber-attacks. The goal is to launch an action based on a voluminous score that either warns of a potential attack (generates risk alerts) or attempts to directly prevent a potential attack (generates risk action).

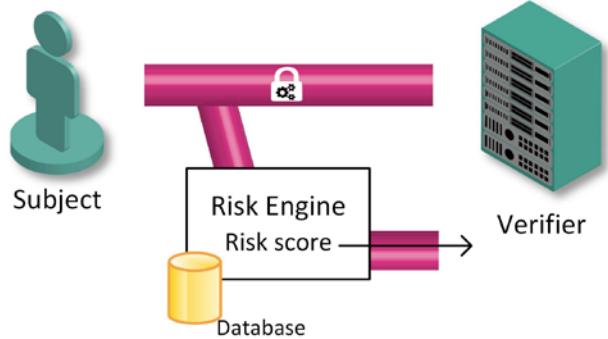


Fig. 3. Risk Based Authentication.

Risk scores can be calculated on the basis of several considerations:

- Risk Engine can accurately monitor the authentication process and detect even minor deviations from this process. These deviations can be caused, for example, by the fact that instead of human authentication the robot (program) trying authentication.
- Comparing the current digital footprint with the history of digital footprint stored in the database. For example, autonomous system (set of IP addresses) from which the user logs in. The version of the software the user uses, etc. This option is being used by the patent [3] and is mentioned in the following text.
- Blacklists.
- Whitelists.

Patent [3] introduces terms:

- “Pre-authentication” as a manner demined both by the identity of the device from which the authentication request originates as well as by available information concerning the identity of the requesting user.
- “Post-authentication” using a user’s transaction history.

Patent [1] gives an interesting example using decision Table 1, table 2 and Table 3. The primary decision table is Table 1 and table 2. Under certain conditions, Table 3 is considered. Score 10 is a likelihood attack, score 0 indicated a low likelihood of attack (fraud).

Patent [1] is tributary for the period in which it was incurred. Currently, users are mainly using mobile applications. For mobile applications, risk-based authentication is even more advantageous. Mobile apps run on a mobile device, so they can read the hardware and software identification of mobile devices and provide risk based authentication. It can therefore provide more information than a web browser.

It is also very important to provide localization data. If the client is authenticated at a short time from two very remote

<sup>1</sup> In the case of the handwriting signature, it is possible to attach a digit, picture, etc. to the signature. This de facto revoke the previous signature without a digit or image.

sites, then it is also poised.

TABLE 1. PRIMARY DEVICE DECISION TABLE

| Period              | Criteria                | Request Attributes   |
|---------------------|-------------------------|--|
| Pre-authentication  | Location information    | <ul style="list-style-type: none"> <li>- City, State, Country information and confidence factors</li> <li>- Connection type</li> <li>- Connection speed</li> <li>- IP address, routing type, and hop times</li> <li>- Internet service provider flag</li> <li>- Autonomous system number</li> <li>- Carrier name</li> <li>- Top-level domain</li> <li>- Second-level domain</li> <li>- Registering organization</li> <li>- A list of anonymizing proxies</li> <li>- Hostnames and routers</li> </ul> |
|                     |                         | X=Missing, M=present and mismatched, * = present and matched   |
|                     |                         | X  |
|                     |                         | M  |
|                     |                         | X/M  |
| Post-authentication | Device information      | <ul style="list-style-type: none"> <li>- Secure Cookies</li> <li>- Flash Cookies</li> <li>- Digitally signed device</li> <li>- Device &amp; display Characteristics:</li> <li>- Operating System characteristics</li> <li>- Browser Characteristics</li> </ul>   |
|                     |                         | T  |
|                     |                         | F  |
|                     |                         | X  |
|                     |                         | T  |
|                     |                         | T  |
|                     |                         | T  |
|                     |                         | T  |
|                     |                         | T  |
|                     |                         | T  |
| Post-authentication | User information        | <ul style="list-style-type: none"> <li>- User identifications</li> <li>- Valid or not valid user</li> <li>- Authentication status</li> </ul>   |
|                     |                         | T  |
|                     |                         | T  |
|                     |                         | T  |
|                     |                         | T  |
|                     |                         | T  |
|                     |                         | T  |
|                     |                         | T  |
|                     |                         | T  |
|                     |                         | T  |
| Post-authentication | Transaction information | <ul style="list-style-type: none"> <li>- Key Value Pairs:</li> <li>- Support multiples</li> <li>- Keys can be defined using Regular Expressions</li> <li>- Values can be defined in ranges</li> <li>- Pages accessed</li> <li>- Time spent on page</li> <li>- Transactions sequences</li> </ul>  |
|                     |                         | T  |
|                     |                         | T  |
|                     |                         | F  |
|                     |                         | X  |
|                     |                         | T  |
|                     |                         | T  |
|                     |                         | T  |
|                     |                         | X  |
|                     |                         | X  |

### III. RISK BASED AUTHENTICATION

The question is how to compare various authentication methods. I propose to use a Risk-based method of comparison. The goal of this method is, if possible, automatically (Online) classify the strength of authentication. The problem is how to quantify the quality of the used authentication. Risk-based authentication uses procedures similar risk analysis. This method starting from empirically prescribed risk level of individual authorization methods. This method does not look too exactly, but during the subsequent evaluation of security incidents ("Feedback") can be risk adjusted, so that over time, this method can be very effective.

TABLE 2. PRIMARY DEVICE DECISION TABLE (SCORE)

| Secure cookie | Flash cookie | Flash data | Browser characteristics | OS characteristics | Score           |
|---------------|--------------|------------|-------------------------|--------------------|-----------------|
| *             | *            | *          | *                       | *                  | 0               |
| X             | *            | *          | *                       | *                  | PATTERN CHECK   |
| M             | *            | *          | *                       | *                  | SECONDARY CHECK |
| X/M           | X            | *          | *                       | *                  | PATTERN CHECK   |
| X/M           | M            | *          | *                       | *                  | SECONDARY CHECK |
| X/M           | X/M          | X          | *                       | *                  | PATTERN CHECK   |
| X/M           | X/M          | M          | *                       | *                  | SECONDARY CHECK |
| X/M           | X/M          | X/M        | M                       | *                  | SECONDARY CHECK |
| X/M           | X/M          | X/M        | X/M                     | M                  | 10              |

TABLE 3. SECONDARY DEVICE DECISION TABLE (SECURE COOKIE MISMATCH)

| Prior Cookie (same device) Browser Characteristics | TRUE Operating System | autonomous system number | Internet service provider | IP stands for IP address | Location | Score |
|--|-----------------------|--------------------------|---------------------------|--------------------------|----------|-------|
| T  | T                     | T                        | T                         | T                        | T        | 0     |
| F  | T                     | T                        | T                         | T                        | T        | 5     |
| X  | T                     | T                        | T                         | T                        | T        | 5     |
| T  | F                     | T                        | T                         | T                        | T        | 10    |
| T  | X                     | T                        | T                         | T                        | T        | 0     |
| T  | T                     | F                        | T                         | T                        | T        | 10    |
| T  | T                     | X                        | T                         | T                        | T        | 0     |
| T  | T                     | T                        | F                         | T                        | T        | 5     |
| T  | T                     | T                        | X                         | T                        | T        | 0     |
| T  | T                     | T                        | T                         | F                        | T        | 5     |
| T  | T                     | T                        | T                         | T                        | F        | 0     |
| T  | T                     | T                        | T                         | X                        | T        | 0     |
| T  | T                     | T                        | T                         | T                        | F        | 5     |
| T  | T                     | T                        | T                         | T                        | X        | 0     |
| F  | F                     | F                        | F                         | F                        | F        | 10    |
| X  | X                     | X                        | X                         | X                        | X        | 10    |

T=TRUE, F=FALSE, X=Missing

### IV. KNOWLEDGE RISK BASED AUTHENTICATION

For the category of authentication as knowledge is practical to use the set of security features. Example of security features in on Table 4.

For each of this security features we express the risk weight  $v_i$  of security features number  $i$ . To avoid big differences in case of adding new security features, we assume:

$$\sum_{i=1}^n v_i = 1$$

For the  $k$ -th authentication mechanism shows the security features as the Risk coefficient  $r_i^k$  gets 1 or 0 depending on whether or not the security features are met. Quality (strength) of authentication mechanism  $k$  can be expressed as:

$$q^k = \sum_{i=1}^n v_i r_i^k$$

In the example given in Table 4, the result is the value is 0.2.

## V. POSSESSION RISK BASED AUTHENTICATION

TABLE 4. EXAMPLE OF SECURITY FEATURES FOR KNOWLEDGE BASE AUTHENTICATION

|    | Security features                   | Classical Password authentication |         |
|----|-------------------------------------|-----------------------------------|---------|
|    |                                     | Weight $v_i$                      | $r_i^k$ |
| 1  | Password change supported           | 1/10                              | 0       |
| 2  | Password reset supported            | 1/10                              | 0       |
| 3  | Password eavesdropping is possible  | 1/10                              | 0       |
| 4  | Password guessing is possible       | 1/10                              | 0       |
| 5  | Password elicitation is possible    | 1/10                              | 0       |
| 6  | Time de-synchronization is possible | 1/10                              | 1       |
| 7  | Anti-desynchronization              | 1/10                              | 0       |
| 8  | Data bearer revocation supported    | 1/10                              | 1       |
| 9  | User anonymity guaranteed           | 1/10                              | 0       |
| 10 | User un-traceability guaranteed     | 1/10                              | 0       |

For the category of authentication categorized as possession we can define the set of security features. For example:

- Cryptographic material does not stored in secured environment (data bearer)
- Access to cryptographic material without password or PIN
- Cryptographic material is exportable
- Cryptographic material does not physically protected against unauthorized access

For each of this security features we express the risk weight  $w_i$ . And similarly assume that:

$$\sum_{i=1}^m w_i = 1$$

For the  $k$ -th authentication mechanism we define the Risk coefficient  $R_i^k$ , Quality (strength) of authentication mechanism  $k$  can be expressed as:

$$Q^k = \sum_{i=1}^m w_i R_i^k$$

## VI. INHERENCE RISK BASED AUTHENTICATION

In this category is traditionally considered biometric characteristics of the person. However, the use of biometric characteristics of persons has many disadvantages. Biometric features cannot be revoked, so have many common features with traditional passwords. In addition, biometric authentication brings complications with the protection of personal data.

In this category of authentication, we will mainly consider digital footprint. We will evaluate the correlation between the information from previous communications and the currently

identified footprint. The correlation coefficient  $\rho$  is from the interval of  $<-1,1>$ . Negative values are important when detecting abnormalities in digital track. With them it is possible e.g. when detecting certain abnormal or decrease the overall weight of authorization. We can use method described in [3] but we need to transform the score to match the  $\rho$  definition domain  $<-1,1>$ .

## VII. OMNIFACTOR AUTHENTICATION

In kind of omnifactor authentication we assume that a user from a set of authentication methods has chosen the method  $k$ . The result quality  $QQ^k$  is weighted sum of individual categories:

$$QQ^k = W_1^k q^k + W_2^k Q^k + W_3^k \rho^k + ext$$

Wight  $W_i^k$  we choose zero in the case that the category is not used and non-zero in case of categories in terms of technology, algorithms and parameters, which ensure increasing quality of authentication. Item  $ext$  mean external authentication, they cannot measure, so we have to determine subjectively.

Using weights Wight  $W_i^k$  can be taken into account independency of various authentication factor.

## VIII. CONCLUSION

It may seem that the problem is to determine the weights  $v_i$  and  $w_i$ . However, at the beginning, it can determine the same weight  $v_i$ , respectively  $w_i$ . Based on the evaluation of security incidents, we can modify individual weight. This will make the model more accurate.

Similarly, we evaluate provided information (services), i.e. assets. If we appreciate an asset, for example, the value of X, then for providing this asset we allow only authentication methods  $k$  with

$$QQ^k \geq X$$

It should also be noted that risk-based authentication could also have drawbacks. Can generate False Positives in the usual cases e.g.: the client purchases a new mobile device; the client will forget to make a payment order prior to the holiday and make it out of an exotic country etc.

## IX. FURTHER WORK

Further work I'll focus on more precise definition of risk weight and simulation of examples. Next problem is re-authentication. It is situation when authenticated does not have sufficient rights and needs to increase score of authentication.

## REFERENCE

- [1] L. Dostalek a J. Ledvina, „Strong Authentication for Mobile Application,“ *International Conference of Applied Electronics, č. IEEE CFP1569A-PRT*, pp. 23-26, September 2015.
- [2] L. Dostálek, „Strong Authentication for Internet Application,“ v *16th European Conference on Cyber Warfare and Security*, Dublin, 2017.
- [3] T. E. Varghese, J. B. Fisher, S. L. Harris a D. D. Boseo, System and Method for Fraud Monitoring, Detection, and Tired User Authentication, Patent N0.: US 7,908,645 B2, 2011.

# The Method of Factorizing Multi-Digit Numbers Based on the Operation of Adding Odd Numbers

Mykhailo Kasianchuk<sup>1</sup>, Igor Yakymenko<sup>1</sup>, Stepan Ivasiev<sup>2</sup>, Ruslan Shevchuk<sup>3</sup>, Lidiya Tymoshenko<sup>4</sup>

1. Department of Computer Engineering, Ternopil National Economic University, UKRAINE, Ternopil, 8 Chekhova str., email: [kasyanchuk@ukr.net](mailto:kasyanchuk@ukr.net), [iyakymenko@ukr.net](mailto:iyakymenko@ukr.net),

2. Department of Cyber Security, Ternopil National Economic University, UKRAINE, Ternopil, 8 Chekhova str., email: [stepan.ivasiev@gmail.com](mailto:stepan.ivasiev@gmail.com)

3. Department of Computer Science, Ternopil National Economic University, UKRAINE, Ternopil, 8 Chekhova str., email: [rsh@tneu.edu.ua](mailto:rsh@tneu.edu.ua)

4. Department of Computer Science and Information Systems Management, Odesa National Polytechnic University, UKRAINE, Odesa, 1 Shevchenko Str., e-mail: [lmt0902@gmail.com](mailto:lmt0902@gmail.com)

**Abstract:** The method for factorizing multi-digit numbers based on the addition of odd numbers is developed in this article. The temporal characteristics of software implementations of the proposed method, the classical Fermat method and its improvement are experimentally investigated. It is revealed that the proposed algorithm, whose time reduction is linear, has an advantage over the other two with a large difference between the multipliers that are factorized. Other dependencies are declining parabolic.

**Keywords:** factorization, Fermat method, addition, multi-digit numbers, time complexity.

## I. INTRODUCTION

The factorization of a natural number or its decomposition into simple multipliers belongs to the main problems of the modern theory of numbers [1-3], various forms of the system of residual classes [4, 5] and plays an important role in cryptographic information security systems [6-8]. As a rule, large-scale computing and time resources are required for the expansion of the multiplicity of the number obtained as a result of the product of two large prime numbers [9].

This circumstance is used in cryptographic algorithms as protection against potential hacking attempts: for the creation and multiplication of two prime numbers of large bits, it does not require significant costs [10], and their factorization is a long and laborious process. Despite the considerable efforts of scientists in this direction, to date, there are no quantum algorithms for factorizing multi-digit numbers, which allow executing the timetable for the multipliers at a practical time. But the proof that there is no solution to this problem in polynomial time is also absent [11].

Particular interest to the factorization problem was the emergence of cryptographic algorithm encryption RSA, which uses its computational complexity [12].

The decomposition of numbers containing in a binary record up to 768 bits inclusive was obtained for polynomial time by the modern methods (in particular, the general method of the numerical field sieve) [13]. However, further increase in the dimension significantly increases the complexity of the scheduling operation of a composite

number. In connection with this, it is necessary to develop other approaches to solving the problem of factorization of integers.

## II. FERMAT'S FACTORIZATION METHOD

Fermat's method is the most widely used method [11], which is based on the search for pairs of natural numbers  $A$  and  $B$ , for which equality  $n=A^2-B^2$  is performed, where  $n=p\cdot q$  is known integer, which is the product of two unknown primes  $p$  and  $q$ , which should be found.

For this purpose  $m=\sqrt{n}$  is searched and then the parameter  $f(x)=(m+x)^2-n$  is calculated, where  $x=1, 2, 3, \dots$ , as long as some value  $f(x)$  will not be equal to the full square of a number, for example  $B^2$ .

Then, respectively,  $A^2=(m+x)^2$  and the desired decomposition will be  $n=p\cdot q=A^2-B^2=(A-B)(A+B)$ .

The most computable complex operations in this case are elevation to a square and the search for a square root. In [14], an improved Fermat's factorization method is proposed, where the condition is used that squares of integers can be represented as a sum of odd numbers, the number of which is equal to this number:

$$s^2 = \sum_{i=1}^s (2i-1) \quad (1)$$

Therefore, having found  $m$  ta  $f_1(x)=f(x)$  when  $x=1$ , the following steps occur according to the expression  $f_i=f_{i-1}+2(m+i)-1$ , where  $i=2, 3, 4, \dots$  as long as  $f_i$  will not be a full square of a certain number.

The decomposition for multipliers will be determined by this expression:  $n=(m+i_f)(m+i+f_i)$ .

Table 1 provides an example of factorization using the classical and improved Fermat's method for  $n=4717$  ( $m_1=\sqrt{4717}=68$ ).

TABLE 1. EXAMPLE OF THE FACTORIZATION OF THE NUMBER 4717 USING CLASSICAL AND IMPROVED FERMAT'S METHOD

| $x$ | $m+x$ | $f(x)$ , classical method | $f(x)$ , improved method |
|-----|-------|---------------------------|--------------------------|
| 1   | 62    | $69^2-4717=44$            | $62^2-4717=44$           |
| 2   | 63    | $70^2-4717=183$           | $44+139=183$             |
| 3   | 64    | $71^2-4717=324=18^2$      | $183+141=324=18^2$       |

Thus the decomposition of the number 4717 on simple multipliers is obtained:  $4717=71^2-18^2=(71+18)(71-18)=89 \cdot 53$ . for  $n=53 \cdot 89=4717$  ( $m_1=\lceil\sqrt{4717}\rceil=68$ ).

It should be noted that the number of iterations in both methods is the same. However, in the improved Fermat's method, the operation of elevating to a square of large numbers is excluded. In addition, arithmetic operations are performed over numbers of a much smaller digit than the classical ones. But the most computationally labor-intensive operation is the extraction of a square root.

Therefore, the purpose of our work is to develop a factorization algorithm based on the Fermat's method, in which there will be no square root search operation, which is especially important for work with numbers of large bit rate for the experimental study, as well as an experimental study of the time characteristics of the software implementation of the Fermat's method, its modification and the proposed algorithm.

### III. DESCRIPTION OF THE PROPOSED FACTORIZATION ALGORITHM

The property (1) is also used in the proposed algorithm, the parameters  $m_1=m$  and  $f_{11}=f_1$  are calculated similarly to the previous case. Then the sequence of operations  $f_{1i}=f_{1 i-1}-r_{1i}$ ,  $r_{11}=1$ ,  $r_{1 i-1}=2i-3=r_{1 i-2}+2$ ,  $i=2, 3, \dots$  are performed until the condition for some  $i$  is not fulfilled:

$$f_{1i}-r_{1i} \leq 0. \quad (2)$$

Further calculations occur in such a way when performing a strict inequality (2):  $m_2=m_1+1$ ;  $f_{21}=f_{11}+2$   $m_2+1-r_{11}$ ,  $r_{21}=r_{11}+2$ . Searching for the following values  $f_{2i}$ ,  $r_{2i}$  is carried out in the same way as the previous case. In the general case, these calculations can be described by the following expressions:

$$f_{ji+1}=f_{ji}-r_{ji} \leq 0, r_{ji+1}=r_{ji}+2, \text{ if } f_{ji}-r_{ji}>0; \quad (3)$$

$$f_{ji+1}=f_{ji}+2m_j+1-r_{ji} \leq 0, r_{ji+1}=r_{ji}+2, \text{ if } f_{ji}-r_{ji}<0. \quad (4)$$

Under the condition  $f_{ji}-r_{ji}=0$ , the desired quantities  $A=(m+j)$  and  $B=\sqrt{(m+j)^2-n}$  are determined which will be a positive integer, the desired quantity, which will be a natural number.

It should be noted that in this case the value of the parameter  $j$  corresponds to the number of steps in the classical and improved Fermat's method.

Fig. 1 shows a block diagram of the developed algorithm, and in the Table 2, the corresponding example is presented

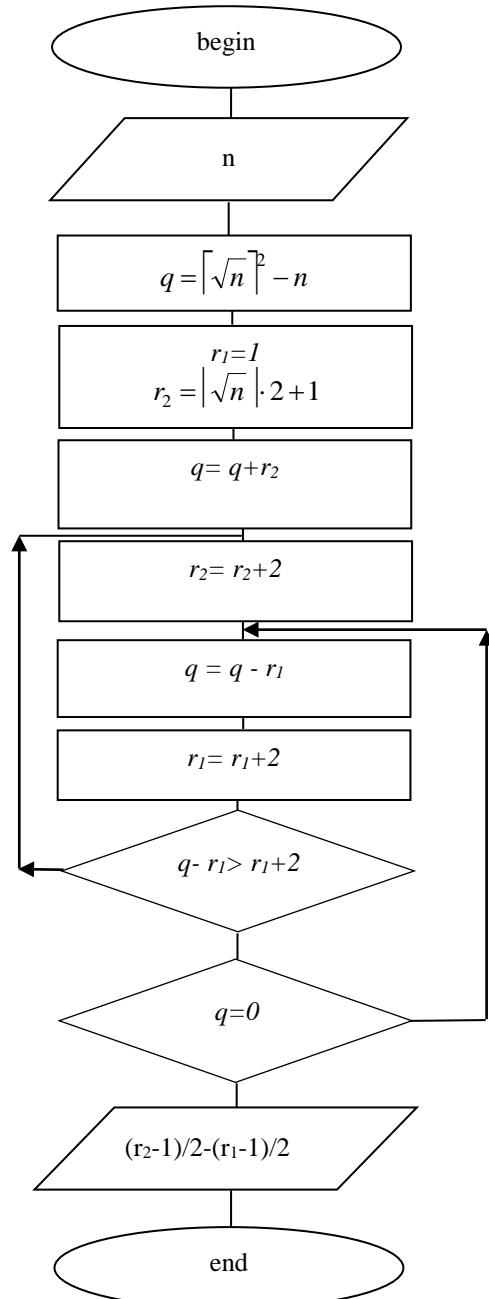


Fig. 1. Block diagram of the developed algorithm

TABLE 2. THE EXAMPLE OF FACTORIZATION OF THE NUMBER 4717 BY THE PROPOSED ALGORITHM

| i | $j=1, m_1=68$  |          | $j=2, m_2=69$  |          | $j=3, m_3=70$   |          |
|---|----------------|----------|----------------|----------|-----------------|----------|
|   | $f_{1i}$       | $r_{1i}$ | $f_{2i}$       | $r_{2i}$ | $f_{3i}$        | $r_{3i}$ |
| 1 | $69^2-4717=44$ | 1        | $8+139-13=134$ | 15       | $14+141-27=128$ | 29       |
| 2 | $44-1=43$      | 3        | $134-15=119$   | 17       | $128-29=99$     | 31       |
| 3 | $43-3=40$      | 5        | $119-17=102$   | 19       | $99-31=68$      | 33       |
| 4 | $40-5=35$      | 7        | $102-19=83$    | 21       | $68-33=35$      | 35       |
| 5 | $35-7=28$      | 9        | $83-21=62$     | 23       | $35-35=0$       |          |
| 6 | $28-9=19$      | 11       | $62-23=39$     | 25       |                 |          |
| 7 | $19-11=8<13$   | 13       | $39-25=14<27$  | 27       |                 |          |

Consequently, the decomposition of the number 4717 into simple multipliers are carried out in this way:  
 $B = \sqrt{(68+3)^2 - 4717} = 18$ ,  $4717 = 71^2 - 18^2 = (71+18) \cdot (71-18) = 89 \cdot 53$ . This procedure is performed without the use of a computationally cumbersome operation extracting the square root. Additionally, addition and subtraction are performed over smaller numbers than in the two previous methods, although the number of these operations is greater.

#### IV. EXPERIMENTAL STUDY OF TIME CHARACTERISTICS OF FACTORIZATION METHODS

The multiplier  $p$  was chosen to be fixed and equal to the largest prime number of a certain bit rate for the experimental study of the factorization of the number  $n=p \cdot q$ .

Then, a number of prime numbers of the same digit are determined that of  $p$ , of which 1000 values for the number  $q$  were selected uniformly in the order of growth.

After multiplying  $p$  by  $q$ , the timer fixed the factorization time for each of the 1000 received products.

The time characteristics of the software implementation of the classical (curve 1) and improved (curve 2) Fermat's methods, as well as the proposed algorithm (curve 3) are presented in Figures 2 and 3 for the bit rates 30 ( $p=1073741789$ ) and 32 ( $p=4294967291$ ) bits respectively.

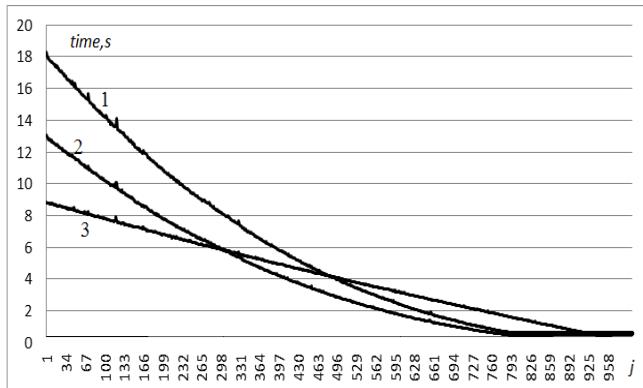


Fig. 2. Time characteristics of software implementation of factorization methods for multipliers with bit rate 30 bits ( $j$  – number in the sample)

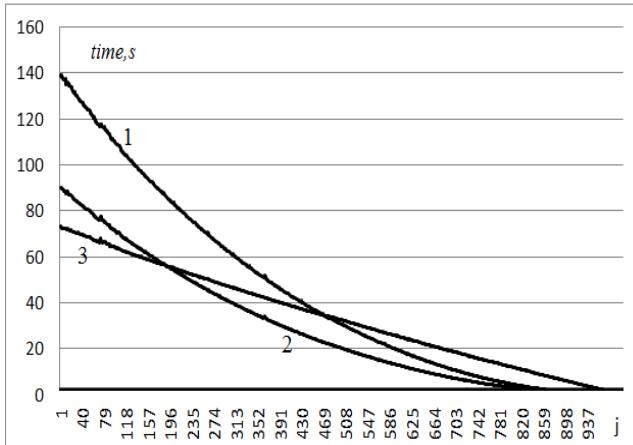


Fig. 3. Time characteristics of software implementation of factorization methods for multipliers with bit rate 32 bits ( $j$  – number in the sample)

All calculations were carried out on a portable computer Lenovo B50-70 with processor Intel Pentium 3558U (1.7 GHz). The amount of RAM in the device was 4 GB. When designing the computing software, a high level programming language C ++ was selected, which allows you to transform the codes into different architectures and operating systems.

All graphics are descending character, indicating a decrease in the factorization time with a decrease in the difference between the multipliers, whose product is factorized. The proposed algorithm, whose time reduction is linear, has an advantage over the other two at small values of  $q$ . With increasing  $q$  the least time is characterized by an improved Fermat's algorithm.

Further growth of  $q$  leads to the fact that the proposed algorithm uses the most time compared to the other two. It should be noted that with increasing the bit rate number of the point of intersection of the straight line with the curves obtained by using the Fermat's method, are shifted to the left on the graphs.

Figure 4 depicts the graphs of the dependence of the average factorization time of three methods on the bit rate number for 1000 values selected by the above-described method. It can be seen that all graphs grow in parabolic law with increasing bit rate. The least average factorization time is characterized by an improved Fermat's algorithm, and the largest one is classical.

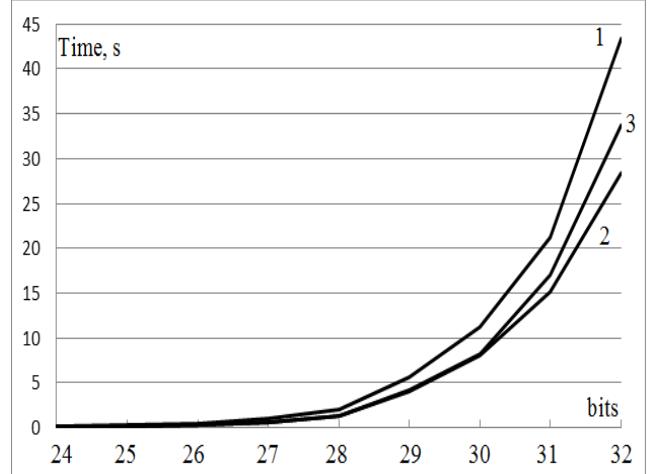


Fig. 4. Graphs of the dependence of the average factorization time on the number of digits: 1 – the classical Fermat's method; 2 – an improved Fermat's method; 3 – the proposed algorithm

Consequently, it is advisable to apply the proposed method when there is a big difference between the numbers whose product is to be factorized.

The Table 3 shows the results of the factorization of the product of two prime numbers for a fixed  $p=3571$  and a variable  $q$  ( $t_1$  – the classical Fermat's method,  $t_2$  – the improved Fermat's method,  $t_3$  – the proposed algorithm).

It can be seen from the Table 3 that the improved Fermat's method increases the speed factorization approximately into 1.1-1.2 times, and the proposed method – into 5-11 times compared with the classical Fermat's method.

TABLE 3. INVESTIGATION OF THE TIME CHARACTERISTICS

| $p$  | $q$        | $p \cdot q$    | $t_1$ , s | $t_2$ , s | $t_3$ , s |
|------|------------|----------------|-----------|-----------|-----------|
| 3571 | 99991      | 357067861      | 0,031     | 0,031     | 0,016     |
| 3571 | 323789     | 1156250519     | 0,172     | 0,157     | 0,032     |
| 3571 | 523771     | 1870386241     | 0,297     | 0,281     | 0,047     |
| 3571 | 723791     | 2584657661     | 0,438     | 0,406     | 0,063     |
| 3571 | 1913803    | 6834190513     | 1,156     | 1,109     | 0,203     |
| 3571 | 2913803    | 10405190513    | 1,781     | 1,703     | 0,313     |
| 3571 | 7913809    | 28260211939    | 4,86      | 4,656     | 0,875     |
| 3571 | 9913807    | 35402204797    | 6,265     | 5,828     | 1,094     |
| 3571 | 19191383   | 68532428693    | 12,843    | 12,156    | 2,407     |
| 3571 | 39192331   | 139955814001   | 26,875    | 26,235    | 4,969     |
| 3571 | 49392341   | 176380049711   | 32,859    | 34,406    | 6,172     |
| 3571 | 139392347  | 497770071137   | 103,078   | 86,391    | 18,031    |
| 3571 | 337392373  | 1204828163983  | 214,141   | 220,375   | 42,483    |
| 3571 | 931392317  | 3326001964007  | 646       | 571,828   | 116,453   |
| 3571 | 1931392319 | 6897001971149  | 1112,657  | 1106,469  | 269,156   |
| 3571 | 2971215073 | 10610209025683 | 2775,718  | 2765,922  | 437,531   |
| 3571 | 6712170737 | 23969161701827 | 11159,56  | 9080,454  | 937,67    |

## V. CONCLUSIONS

This paper is devoted to the experimental study of the time complexity for the factorization of many digit numbers using the classical and improved Fermat method, and also the factorization algorithm based on the subtraction operation is proposed and investigated. Appropriate graphic dependencies have been constructed. It is shown that the proposed algorithm is characterized by less temporal complexity in the case where the big difference between the numbers whose product must be factorized.

## REFERENCES

- [1] D. Venturi "Lecture Notes on Algorithmic Number Theory" *New-York-Berlin: Springer-Verlag*, 2009. 217 p.
- [2] S. Ishmukhametov, A. Boyko and D. Ziyatdinov "On an approach to the problem of the factorization of natural numbers", *Proceedings of High Schools. Mathematics*, 2011, pp. 15–22.
- [3] V. Shoup "Computational Introduction to Number Theory and Algebra", *Cambridge University Press, Sec.Edition*, 2005, 600 p.
- [4] M. Kasianchuk, "The Construction of the modified Perfect Form of Residual Classes System Using Factorization", *Radio Electronics, Computer Science, Control*, 2017, Vol.42, №3, pp. 53-59.
- [5] Ya.M. Nykolaychuk, M.M. Kasianchuk and I.Z. Yakymenko "Theoretical Foundations of the Modified Perfect Form of Residue Number System", *Cybernetics and Systems Analysis*, 2016, V.52, №2, pp. 219-223.
- [6] A.V. Agranovsky and R.A. Hadi "Practical cryptography: algorithms and their programming", *Moscow: Solon-Press*, 2009, 256 p.
- [7] O.N. Vasilenko "Theoretical and numerical algorithms in cryptography", *Moscow: Center for Continuous Mathematical Education*, 2003, 326 p.
- [8] N. Koblitz "Course of Number Theory and Cryptography", *Moscow: TVP*, 2001, 260 p.
- [9] R. Crandall, C. Pomerance "Prime Numbers: A Computational Perspective. Chapter 5: Exponential Factoring Algorithms", *Springer-Verlag: New York*, 2005, pp. 2-6.
- [10] D. Kozaczko, S. Ivasiev, I. Yakymenko and M. Kasianchuk "Vector Module Exponential in the Remaining Classes System", *Proceedings of the 2015 IEEE 8th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS'2015)*, Warsaw, Poland, 2015, pp. 161-163.
- [11] S. Ishmukhametov "Methods for the factorization of natural numbers: a tutorial", *Kazan: Kazan University*, 2011, 190 p.
- [12] E.B. Makhovenko "Theoretical and numerical methods in cryptography: Textbook", *Moscow: Helios ARV*, 2006, 320 p.
- [13] RSA Challenge. URL: <https://www.emc.com/emc-%20plus/rsa-labs/historical/the-rsa-challenge-numbers.htm>.
- [14] M. Karpiński, S. Ivasiev, I. Yakymenko, M. Kasianchuk and T. Gancarczyk "Advanced method of factorization of multi-bit numbers based on Fermat's theorem in the system of residual classes", *Proc. of 16th International Conference on Control, Automation and Systems (ICCAS-2016)*, Gyeongju, Korea, V.1, October, 2016, pp. 1484–1486.

# Hardware Components for Post-Quantum Elliptic Curves Cryptography

Rodrigue Elias<sup>1</sup>, Valerii Hlukhov<sup>2</sup>, Mohammed Rahma<sup>2</sup>, Ivan Zholubak<sup>2</sup>

1. School of Engineering at Lebanese International University, LEBANON, Beirut, 2<sup>nd</sup> floor - Block F- Beirut Campus, P.O. Box: 146404, Mazraa, e-mail: [rodrigue.elias@liu.edu.lb](mailto:rodrigue.elias@liu.edu.lb)

2. Computers Dept., Lviv Polytechnic National University, UKRAINE, Lviv, 12 Bandera Str., e-mail: [gulukhov@polynet.lviv.ua](mailto:gulukhov@polynet.lviv.ua)

**Abstract:** Investigations in the sphere of quantum calculations form up new challenges in the public key cryptography. Currently known public key crypto algorithms may be compromised with the implementation of quantum computers. The workgroups of ETSI and NIST determined the promising trends, within the framework of which there could be obtained acceptable solutions and one of the trends is the use of algorithms that process points of supersingular elliptic curves. Hardware implementations of components for processing points of elliptic curves are well known. The purpose of this publication is to investigate how they can be used to process points of supersingular elliptic curves.

**Keywords:** hardware components, post-quantum, elliptic curves, cryptography.

## I. INTRODUCTION

The advent of large-scale quantum computing offers great promise to science and society, but brings with it a significant threat to global information infrastructure. Public-key cryptography - widely used on the internet today – relies upon mathematical problems that are believed to be difficult to solve given the computational power available now and in the medium term.

However, popular cryptographic schemes based on these hard problems – including Elliptic Curve Cryptography – will be easily broken by a quantum computer. This will rapidly accelerate the obsolescence of currently deployed security systems and will have dramatic impacts on any industry where information needs to be kept secure.

Quantum-safe cryptography refers to efforts to identify algorithms that are resistant to attacks by both classical and quantum computers, to keep information assets secure even after a large-scale quantum computer has been built [1].

Supersingular isogeny Diffie–Hellman key exchange (SIDH) is a post-quantum cryptographic algorithm used to establish a secret key between two parties over an otherwise insecure communications channel. It is analogous to the Diffie–Hellman key exchange, but is designed to resist cryptanalytic attack by an adversary in possession of a quantum computer.

## II. THE SUPERSINGULAR ISOGENY DIFFIE–HELLMAN BACKGROUND

The supersingular isogeny Diffie–Hellman (SIDH) method works with the set of supersingular elliptic curves  $E$  over

Galois Field  $GF(p^2)$ . An isogeny of an elliptic curve  $E$  is a rational map from  $E$  to another elliptic curve  $E'$  which is also a group homomorphism. Provided the isogenies are separable, they are determined by the points inside their kernel up to isomorphisms of  $E'$ .

The SIDH method works with a prime of the form  $p = (w_A)^{e_A} (w_B)^{e_B} (f) \pm 1$  where  $w_A$  and  $w_B$  are small primes and an elliptic curve  $E$  defined by the equation:  $y^2 = x^3 + ax + b$ . SIDH builds an isogeny map from a single elliptic curve point which is taken as the generator for the isogeny's kernel. This point is chosen to be a random linear combination to two fixed points chosen to be in the kernel of the isogeny.

The  $j$ -invariant of an elliptic curve  $E$  is a fixed function of a set of isomorphic curves. It is computed from the parameters that define the curve. For an elliptic curve  $E$  defined by the equation:  $y^2 = x^3 + ax + b$  the  $j$ -invariant of the curve  $E$  is  $j(E) = 1728 \frac{4a^3}{4a^3 + 27b^2}$ .

The security of SIDH is closely related to the problem of finding the isogeny mapping between two supersingular elliptic curves with the same number of points. In [3] it was shown that the security of SIDH will be  $O(p^{1/4})$  for classical computers and  $O(p^{1/6})$  for quantum computers. This suggests that SIDH with a 768-bit prime ( $p$ ) will have a 128-bit security level.

In 2014, researchers at the University of Waterloo developed a software implementation of SIDH. They ran their partially optimized code on an x86-64 processor running at 2.4 GHz. For a 768-bit modulus they were able to complete the key exchange computations in 200 milliseconds thus demonstrating that the SIDH is computationally practical [4].

In 2016, researchers from Microsoft posted software for the SIDH which runs in constant time (thus protecting against timing attacks) and is the most efficient implementation to date [5].

In 2017, researchers from Florida Atlantic University developed the first FPGA implementations of SIDH for 83-bit and 124-bit quantum security levels [6].

## III. DEFINITION OF ISOGENY

For supersingular elliptic curves there is well known Vélu algorithm [7] for isogeny [8]: Let  $E_1$  and  $E_2$  be elliptic curves over the field  $F$ . The isogeny  $E_1 \rightarrow E_2$  over  $F$  is a non-

constant rational mapping over  $F$ , which is also a group homomorphism

$$(x, y) \rightarrow \left( \frac{f_1(x, y)}{f_2(x, y)}, \frac{g_1(x, y)}{g_2(x, y)} \right), \text{ where } f_1, g_1, f_2, g_2 \text{ are polynomials.}$$

For example,  $F = GF(19)$ ,

$$\text{elliptic curve E1: } y^2 = x^3 + x + 1;$$

$$\text{elliptic curve E2: } y^2 = x^3 + 4x + 13$$

$$\#E1 = \#E2 = 21;$$

$$(x, y) \rightarrow \left( \frac{x^3 - 4x^2 - 8x - 8}{x^2 - 4x - 4}, \frac{x^3 y - 6x^2 y - 5xy - 6y}{x^3 - 6x^2 - 7x - 8} \right).$$

So, to determine the isogeny, it is necessary to perform the operations of addition, multiplication and inverse element calculation in the Galois field  $GF(p^2)$ .

#### IV. ESTIMATION OF THE SOFTWARE TIME COMPLEXITY OF OPERATIONS IN THE GALOIS FIELDS

In works [11] and [12], the evaluation of the ability of data protection means to counteract attacks by hackers was carried out. The definition of Galois field in which hackers work

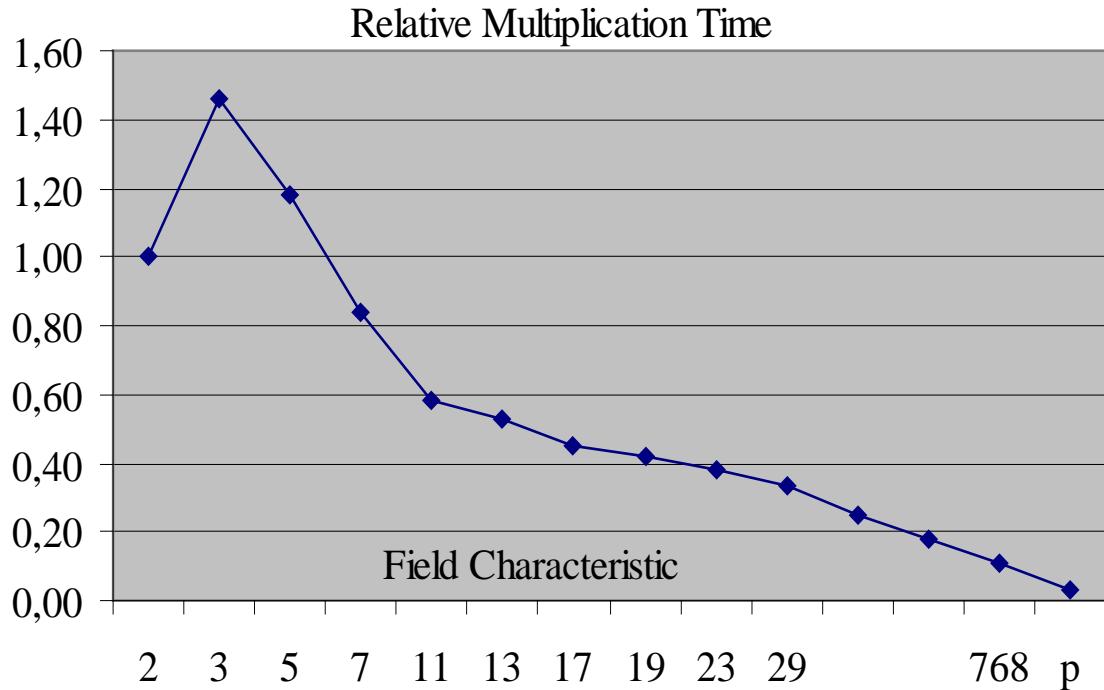


Fig. 1. Relative time complexity of software multiplying in a different fields.

The relative time complexity was determined by the ratio of the multiplication time in the field  $GF(d^m)$  to the multiplication time in the field  $GF(2^{99})$ .

As can be seen from the Table 1, software multiplication of triple extended field elements has the longest execution time. It provides hardware cryptoprocessors based on such fields additional protection against hacking. Software-implemented operations on simple field elements are executed in the fastest way, that indicates the inappropriateness of cryptographic processors based on such fields. Multiplication in binary fields has one of the highest time complexity, it is

hardest was the purpose of the study. It was assumed that hackers use software methods.

One of the computer system hacking methods is the brute-force method [8], in which the general-purpose computer selects all sorts of keys or passwords until one of them fits. The same operations over Galois fields elements are performed both during the execution of the hack program and in the hardware crypto processors. For general purpose computers, it is possible to estimate the time of execution of the main operation, multiplication of the Galois fields elements, for extended fields with different characteristics, but with approximately the same order. The basis for such a check you can take the field  $GF(2^{99})$ . The calculations you can make using the Maple 2017 package [9]. The relative times of execution of such number of multiplications with respect to the time of execution of the same number of operations in the binary field  $GF(2^{99})$  are shown Table 1 and in the Fig. 1 where prime  $p \approx 2^{99}$  is characteristic of prime  $GF(p)$ .

third after multiplication complexities in fields with characteristics 3 and 5. Therefore, the following study will focus on binary fields.

Software multiplication in fields with characteristic 768 which is used in [3] is performed for 10 times faster than in binary fields. Therefore, this system can be hacked using classic computers faster than the system that uses binary fields.

TABLE 1. THE RELATIVE TIME COMPLEXITY OF MULTIPLICATION IN EXTENDED FIELD OF DIFFERENT CHARACTERISTICS

| Field Characteristic | Relative Time |
|----------------------|---------------|
| 2                    | 1,00          |
| 3                    | 1,46          |
| 5                    | 1,18          |
| 7                    | 0,84          |
| 11                   | 0,59          |
| 13                   | 0,53          |
| 17                   | 0,45          |
| 19                   | 0,42          |
| 23                   | 0,38          |
| 29                   | 0,33          |
| ...                  | ...           |
| 768                  | 0,11          |
| p                    | 0,03          |

## V. ESTIMATION OF THE HARDWARE TIME COMPLEXITY OF OPERATIONS IN THE GALOIS FIELDS

Implemented in modern FPGA hardware multipliers for extended Galois field GF ( $d^m$ ) with approximately the same number of elements  $d^m \approx 2^n$  were estimated in [13] in terms of their time complexity to determine the fields in which the multiplier will have the least time complexity. Relative to GF( $2^n$ ) results of estimation is shown in Fig. 2.

## VI. ESTIMATION OF THE SOFTWARE-HARDWARE TIME COMPLEXITY OF OPERATIONS IN THE GALOIS FIELDS

We will assume that the hardware implementation is used by users of data protection tools, and software is used by hackers. Then we can introduce a generalized time complexity index. We will calculate this indicator as the ratio of software time complexity to hardware time complexity for the same fields. When the value of this indicator is greater, then hackers will have more problems, so data protection will be better.

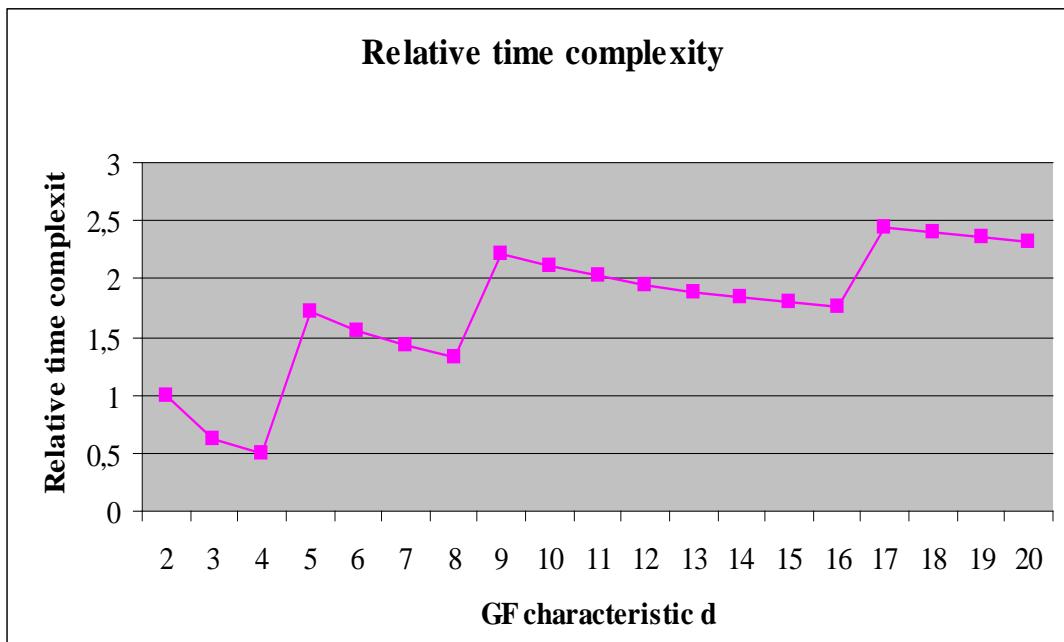


Fig. 2. Relative time complexity of hardware multipliers in different fields.

Results of hardware-software complexity estimation is shown in Fig. 3.

As can be seen from Fig. 3, the trial and binary Galois fields provide the best data protection. Fields with large characteristics provide weaker protection. The use of hardware tools for working in such fields has less effect than for working in binary and ternary fields. As result the use of

hardware tools to work in the field with the characteristic 768 [3] also has less effect than for working in binary and ternary fields.

The use of isogenies of supersingular elliptic curves is oriented toward usage of Galois field with big characteristics, so they are focused on software implementation.

## Relative HS-Complexity

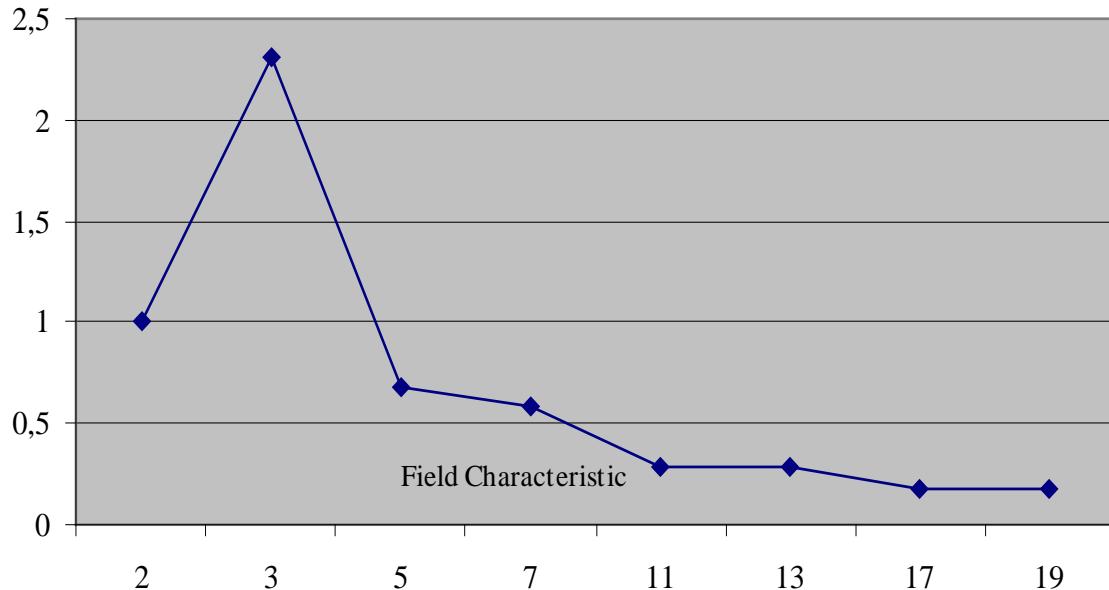


Fig. 3. Relative hardware-software time complexity of multipliers in different fields.

### V. CONCLUSION

The use of isogenies of supersingular elliptic curves is oriented toward software implementation of the method. Hardware implementation of this method will not provide such a reduction in time complexity and increase the degree of data protection as it provides in methods oriented on binary fields.

### REFERENCES

- [1] Quantum-safe cryptography. <http://www.etsi.org/technologies-clusters/technologies/quantum-safe-cryptography>
- [2] Supersingular isogeny key exchange. [http://https://en.wikipedia.org/wiki/Supersingular\\_isogeny\\_key\\_exchange](https://en.wikipedia.org/wiki/Supersingular_isogeny_key_exchange)
- [3] De Feo, Luca; Jao, Plut. "Towards quantum-resistant cryptosystems from supersingular elliptic curve isogenies" (PDF). PQCrypto 2011. Springer. Retrieved 4 May 2014.
- [4] Fishein, Dieter (30 April 2014). "Machine-Level Software Optimization of Cryptographic Protocols". University of Waterloo Library - Electronic Theses. University of Waterloo. Retrieved 21 June 2014.
- [5] Costello, Craig; Longa, Patrick; Naehrig, Michael (2016-01-01). "Efficient algorithms for supersingular isogeny Diffie-Hellman"
- [6] Koziel, Brian; Kermani, Mehran; Azarderakhsh, Reza (2016-11-07). "Fast Hardware Architectures for Supersingular Isogeny Diffie-Hellman Key Exchange on FPGA"
- [7] Jean Vélu. Isogénies entre courbes elliptiques. Comptes Rendus de l'Académie des Sciences de Paris, 273:238–241, 1971.
- [8] SIKE protocol and its stability to classical and quantum attacks. [https://www.ruscrypto.ru/resource/archive/rc2018/files/02\\_Taraskin.pdf](https://www.ruscrypto.ru/resource/archive/rc2018/files/02_Taraskin.pdf)
- [9] Password cracking. [https://en.wikipedia.org/wiki/Password\\_cracking](https://en.wikipedia.org/wiki/Password_cracking).
- [10] Maple User Manual. Copyright © Maplesoft, a division of Waterloo Maple Inc. 2017.
- [11] Hlukhov, V., Zholubak, I., Kostyk, A., Rahma M. (2017), Galois Fields Elements Processing Units for Cryptographic Data Protection in Cyber-Physical Systems. Advances in Cyber-Physical Systems. Volume II, Number 2, 2017. © Lviv Polytechnic National University, pp. 9-18, in press.
- [12] Rodrigue Elias, Valerii Hlukhov, Mohammed Rahma, Ivan Zholubak: FPGA cores for fast multiplicative inverse calculation in Galois Fields. 9th International IEEE Conference Dependable Systems, Services and Technologies DESSERT'2018 UKRAINE, KYIV, MAY 24-27, 2018, in press.
- [13] R. Elias, M. Rahma, V. Hlukhov. Multipliers for Galois fields Time Complexity. ELECTROTECHNIC AND COMPUTER SYSTEMS. Science and Technical. Odessa, No. 22 (98) 2016. Pp. 323-327 (In Ukrainian)

# New Text Encryption Method Based on Hidden Encrypted Symmetric Key

Seddeq E. Ghrare, Haneen A. Barghi, Nora R. Madi

Department of Electrical and Computer Engineering, Faculty of Engineering, University of Gharyan,  
P.O Box 64418 Gharyan - Libya  
email: [eng.ukm@gmail.com](mailto:eng.ukm@gmail.com), [seddeq@jgu.edu.ly](mailto:seddeq@jgu.edu.ly)

**Abstract:** Cryptography is classified into two main categories which are Symmetric Key Cryptography and Asymmetric Key Cryptography. In both categories, the security level provided by any cryptographic algorithm depends on its encryption and decryption keys. In this paper a new encryption and decryption algorithm based on Hidden Encrypted Symmetric Key (HESK) is designed and implemented. The strength of this algorithm is represented in the key used for encryption and decryption process. The key itself is encrypted prior to be used for plain text encryption and cipher text decryption processes, then it is hidden inside the cipher text. The aim of hiding the key is to overcome the problem of distributing the secret key and to make the proposed algorithm more secure and difficult or even impossible to be broken. The proposed algorithm is tested on a set of plain texts of various sizes. The experimental result has been demonstrated that it is difficult to factorize the used key. The main two advantages of the proposed method are represented in the computation simplicity and security efficiency.

**Keywords:** *cryptography; encryption; decryption; plaintext; ciphertext*

## I. INTRODUCTION

On the Internet, information passes from one computer to another through numerous systems before it reaches its destination. Some information, such as banking, electronic payment and electronic voting are very sensitive, therefore it should run and exchanged over the network in a robust manner and safely [1]. Cryptography is considered one of the most used ways to protect the sensitive information and prevent unauthorized people from altering that information.

Cryptography is the science of using mathematics to encrypt and decrypt data. Cryptography enables you to store sensitive information or transmit it across insecure networks (like the Internet) so that it cannot be read by anyone except the intended recipient. While cryptography is the science of securing data, cryptanalysis is the science of analyzing and breaking secure communication. Cryptanalysts are also called attackers. Cryptology embraces both cryptography and cryptanalysis [4].

The history of cryptography can be broadly divided into three phases:

1. From ancient civilizations to the nineteenth century and the first part of the twentieth century, with relatively simple algorithms that was designed and implemented by hand.

2. Extensive use of encrypting electro-mechanical machines, around the period of the Second World War.

3. Ever more pervasive use of computers, about in the last fifty years, supported by solid Mathematical basis.

Cryptography was already used in ancient times, essentially in three kinds of contexts:

- a) Private communications
- b) Art and religion
- c) Military and diplomatic use

A cryptographic algorithm is a function used for both encryption and decryption processes. This function is dependent on a key value necessary for both encryption and decryption [2]. The problem associated with the cryptographic algorithms is the security that can be provided. The strength of any cryptographic algorithm depends on the strength of the keys used. In other words, the problem of low level security of any algorithm arises from the weak encryption and decryption keys that have been used and because of the rapid growth in factorization algorithms; weak encryption and decryption keys were easily factored and discovered. To overcome this problem and to provide a good level security, the used keys should be powerful enough [3].

Based on the used keys, cryptographic algorithms can be classified into two main categories which are asymmetric key cryptographic algorithms and symmetric key cryptographic algorithms [4].

In the first category algorithms, the key used for decryption process is different from the one used for encryption process. It is extremely difficult to determine one key by analyzing the other. This allows for the free distribution of one key (i.e., public), while the key used for decryption is kept private [3,4].

The opposite is true for the second category algorithms, keys used for encryption and decryption processes are the same. This requires that sender and receiver agree on the key prior to any information exchange [3,4]. Both asymmetric and symmetric key cryptography are illustrated in Figs. 1 and 2.

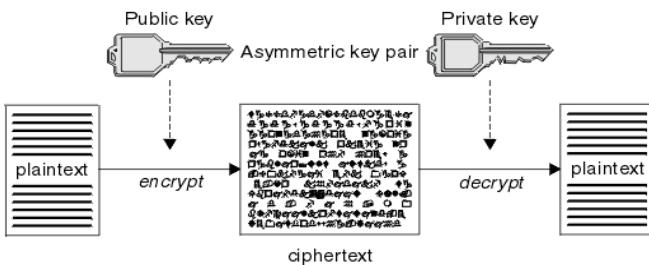


Fig. 1. Asymmetric Key Cryptography.

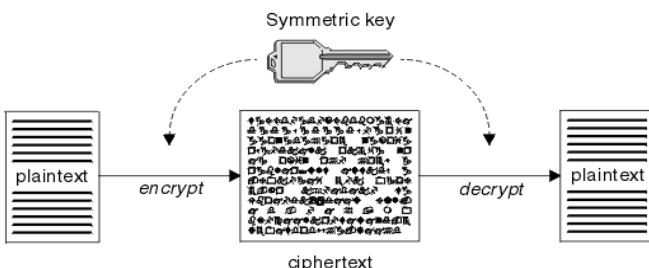


Fig. 2. Symmetric Key Cryptography.

In this paper a new encryption and decryption method based on Hidden Encrypted Symmetric Key (HESK) is designed and implemented. This algorithm starts with reading the plaintext. Then it generates the encryption and decryption key from the plain text. The key itself is encrypted prior to be used for plain text encryption and cipher text decryption processes. Then the encrypted key is hidden in the cipher text. Finally, both the encrypted key and cipher text are sent. The aim of hiding the key in the cipher text is to overcome the problem of distributing the secret key and to make the proposed algorithm more secure and difficult or even impossible to be broken. The proposed method is tested on a set of plain texts of various sizes. The experimental result has been demonstrated that it is difficult to factorize the used key. The main two advantages of the proposed method are represented in the computation simplicity and security efficiency.

## II. METHODOLOGY

The cryptosystem of the proposed algorithm has been divided into three modules, as indicated in Figure 3, which are:

- Key generation module
- Data encryption module
- Data decryption module

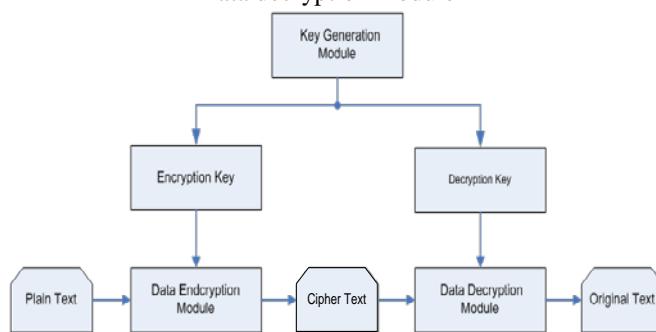


Fig. 3. Cryptosystem of the Proposed Algorithms.

### A. Key Generation Module

This stage involves the generation of two keys, which could be used to encrypt the input data or message to be transferred and to decrypt the received encrypted message at the destination; those two keys are called key1 and key2 respectively. Only one of these keys can be used. In order to increase the security level of the proposed method, the chosen encryption key is encrypted and then used to encrypt the plain text in the following next stage (module).

### B. Data Encryption Module

In this module any data or plain text to be sent to the receiver is encrypted prior to being transferred using the generated and encrypted keys; Key1 or Key2. The used encryption key is then inserted and hidden in the cipher text. Finally, the cipher text, which contains on the encrypted encryption keys, is sent to the destination.

### C. Data Decryption Module

When the encrypted data (**Cipher text**) reaches the receiver, it cannot be read. In order to be read, the hidden encrypted decryption key should be extracted from the cipher text. Then the cipher text is decrypted and converted to its original form (**plain text**) using the extracted key.

The main steps of the proposed algorithm are as follows:

**Step 1:** Read the plain text

**Step 2:** Divide the plain text into two halves

**Step 3:** Generate the encryption and decryption keys

3.1: Key1 = LH (Lower half of the plaintext)

3.2: Key2 = UH (Upper half of the plaintext)

**Step 4:** Encrypt the encryption and decryption key

**Step 5:** Use the resulted encrypted key to encrypt the whole plain text

**Step 6:** Hide the encryption key in the encrypted text (Cipher text)

**Step 7:** Send the cipher text with the hidden key to the intended receiver.

In the previous third step, which involves the key generation, there are two generated keys (key1 and key2) which means that either one can be used. The size of the used key is equal to the half of the plain text size.

The generated key is firstly encrypted, then the encrypted key is used to encrypt the plain text, and finally the encrypted key is hidden inside the cipher text.

At the receiver side the following steps should be followed in order to get the original plain text.

**Step 8:** Extract the hidden key from the cipher text

**Step 9:** Decrypt the cipher text using the extracted key

**Step 10:** Decrypt the key

The encryption and decryption processes are performed using the following equations:

$$C = E(k, p)$$

$$E(k, p) = (p+k) \bmod 26 \quad (1)$$

where :

*C* is the cipher text

*K* is the encryption key

*P* is the plain text

*E* is the encryption algorithm performed to encrypt the plain text (*p*) using the encryption key (*k*)

$$\begin{aligned} P &= D(k, C) \\ D(k, C) &= (p - k) \bmod 26 \end{aligned} \quad (2)$$

where :

*C* is the cipher text  
*K* is the decryption key  
*P* is the plain text

*D* is the decryption algorithm performed to decrypt the cipher text (*c*) using the decryption key (*k*)

The above steps of the proposed algorithm are implemented using Java programming language [5] and tested on a set of text file of different sizes Figure 4 shows the block diagram of the proposed algorithm.

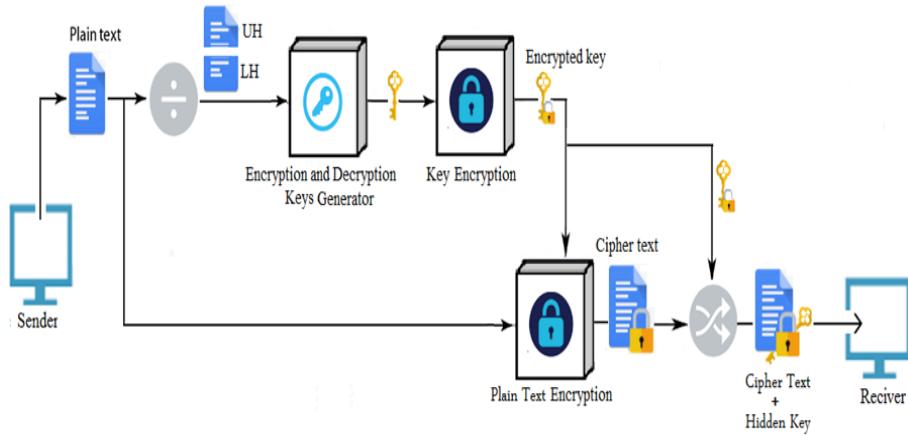


Fig. 4. Block Diagram of the Proposed Algorithm.

### III. RESULTS PRESENTATION

As it was mentioned, the proposed algorithm is implemented using Java programming language. The results obtained by performing the proposed algorithm on the same files of sizes and executed using hardware with the following specifications:

- Windows 7 Ultimate Operating System
- Intel Core3 Processor

- CPU speed of 2.10 GHz
- RAM size of 2 GB
- HDD of 500 GB

The size of the used data files are 1KB, 10KB, 100KB, 300kB and 0.5 MB. The Graphical User Interface (GUI); which is generated by the designed program; is shown in Figure 5.

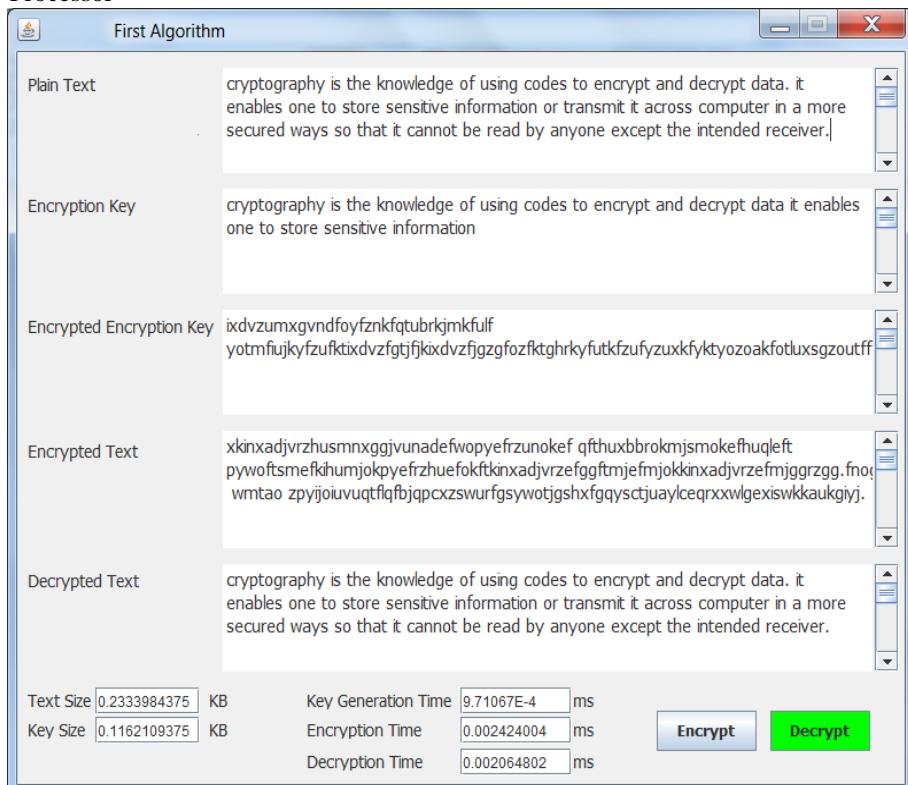


Fig. 5. The GUI of the Proposed Algorithm.

Table 1 shows the time taken for the key generation process, encryption process, and decryption process respectively. Note that when the size of the plain text is increased, the size of the generated encryption key will be longer and will take longer time. Consequently, the generated key will be stronger and harder to be broken or discovered. Moreover, the generated encryption key is encrypted before it was used for encrypting the plain text, then it is inserted into the cipher text to be hidden. So, all those steps were taken in the account and included in the time calculations.

TABLE 1. EXECUTION TIME OF THE PROPOSED ALGORITHM

| Plain Text Size(KB) | Key Generation Time (Sec) | Encryption Time (Sec) | Decryption Time (Sec) |
|---------------------|---------------------------|-----------------------|-----------------------|
| 1                   | 0.80                      | 1.20                  | 1.70                  |
| 10                  | 1.45                      | 2.90                  | 4.54                  |
| 100                 | 5.58                      | 13.87                 | 19.67                 |
| 300                 | 12.62                     | 27.50                 | 31.90                 |
| 500                 | 23.35                     | 64.95                 | 72.10                 |

From the above table, the following comments can be extracted:

The overall computation time taken by the proposed algorithm is 1.20 sec, 3.00 sec, 13.00 sec, 24.00 sec, and 53.00 sec, to encrypt and decrypt plain texts of 1KB, 10KB, 100KB, 300KB, and 500KB respectively including the time taken for key generation and key encryption and hiding..

The time taken for the three stages (key generation, encryption and decryption processes) using the proposed algorithm increases whenever the size of the plain text is increased. The reason as it was explained earlier, because of that whenever the size of plain text is increased; the size of encryption key will be longer and as a result the longer time will be taken for the generation process. This can be seen from the following figure.

From figure 6, it can be noticed that the most of time is elapsed in the decryption process followed by the encryption process then key generation process. Moreover, the time is increasing and getting longer whenever the size of the plain text is increased for all the three processes.

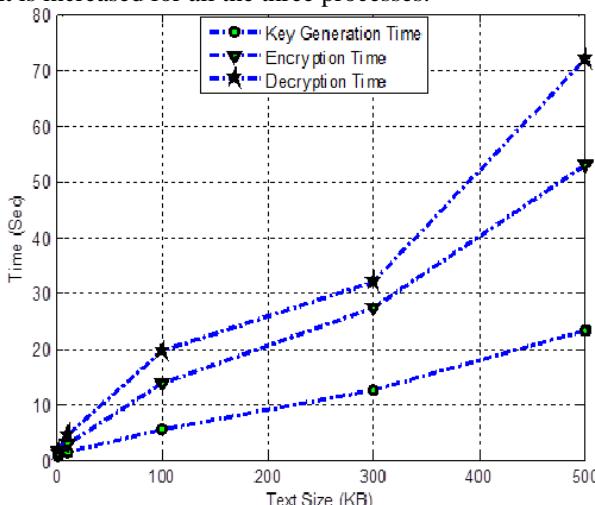


Fig. 6. Performance of the Proposed Algorithm.

#### IV. ANALYSIS OF THE PROPOSED ALGORITHM

For the analysis purpose, the proposed system has been implemented in JAVA programming language for demonstration intention. Proposed system has been analyzed for identification of keyword, identification of keyword distance, identification of polynomial and identification of key stream. This analysis was carried out using both Frequency analysis test and Kasiski analysis test.

##### A. Frequency Analysis Test

Frequency analysis is the study of letters or groups of letters contained in a ciphertext in an attempt to partially reveal the message. The English language (as well as most other languages) has certain letters and groups of letters appear in varying frequencies

In English language, "E" is the most common letter, appearing about 12% of the time (that is just over one in ten letters is an "E"). The next most common letter is "T" at 9%. The full frequency list is given by the graph illustrated in figure 7.

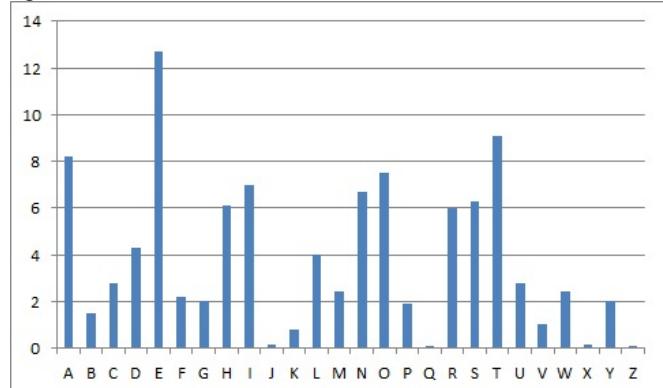


Fig. 7. Frequency of English Letters.

Further frequency analysis is applied for the ciphered text shown in Figure 10. This cipher text in resulted from the plain text enciphered by the encrypted key shown in Figure 9 using the proposed algorithm.

Cryptography is the knowledge of using codes to encrypt and decrypt data. It enables one to store sensitive information or transmit it across computer in a more secured ways so that it cannot be read by anyone excepted the intended receiver.

Fig. 8. Plain Text Sample.

ixdvzumxgvndfoyfznkfqtubrkjmkful  
yotmfiujkyzfuzktixdzfgtjfkixdvzfjgzgfozfk  
tghrkfyutkfzufyzuxkfyktyozoakfotluxsgzoutf  
f

Fig. 9. Generated Encrypted Key.

xkinxadgyzhusmnxxggvunadefwopyefrzunokefqfthu  
xbrokmjsmokefhuqlefptwoftsmefkihumjokpyefrz  
huefokftkinxadjvrzefgftmjefmjokkinxadjvrzefmjg  
grzgg.fnogwmtaozpyijoiuvuqtflqfbjqpcxzswurfgsy  
wotjgshxfqysctjuaylceqrxxwlgexiswkkaukgiy.

Fig. 10. Cipher Text with a Hidden Encryption Key.

The results obtained from applying the frequency analysis test are shown in Table 2.

TABLE 2. COMPARISON RESULTS USING FREQUENCY ANALYSIS

| English alphabet | Frequency of English letters | Proposed cipher |
|------------------|------------------------------|-----------------|
| A                | 8.17                         | 3.00            |
| B                | 1.49                         | 1.30            |
| C                | 2.78                         | 1.30            |
| D                | 4.25                         | 1.75            |
| E                | 12.70                        | 5.70            |
| F                | 2.23                         | 8.80            |
| G                | 2.02                         | 6.55            |
| H                | 6.09                         | 2.60            |
| I                | 6.97                         | 3.50            |
| J                | 0.15                         | 5.25            |
| K                | 0.77                         | 4.80            |
| L                | 4.03                         | 1.75            |
| M                | 2.41                         | 3.90            |
| N                | 6.75                         | 3.00            |
| O                | 7.51                         | 5.25            |
| P                | 1.93                         | 3.00            |
| Q                | 0.10                         | 2.60            |
| R                | 5.99                         | 3.00            |
| S                | 6.33                         | 3.00            |
| T                | 9.06                         | 3.90            |
| U                | 2.76                         | 4.80            |
| V                | 0.98                         | 2.20            |
| W                | 2.36                         | 2.60            |
| X                | 0.15                         | 4.80            |
| Y                | 1.97                         | 3.00            |
| Z                | 0.07                         | 3.50            |

From the above table it can be seen that the frequency of the letters of cipher text obtained by using the proposed algorithm is totally different from the original frequency of the English language letters which proves that the keyword of the proposed algorithm is hard to be revealed using frequency analysis.

#### B. Kasiski Analysis Test

The Kasiski analysis test involves looking for strings of characters that are repeated in the ciphertext. The strings should be three characters long or more for the examination to be successful. The reason this test works is that if a

repeated string occurs in the plaintext, and the distance (period) between corresponding characters is a multiple of the keyword length, the keyword letters will line up in the same way with both occurrences of the string. Then, the distances between consecutive occurrences of the strings are likely to be multiples of the length of the keyword. Thus finding more repeated strings narrows down the possible lengths of the keyword.

The Kasiski test has been applied to the text which is enciphered using the proposed algorithm. The most repeated strings of character and their distances are listed in Table

TABLE 3. REPEATED STRINGS AND PERIODS

| Sequence | zhu | ywo | zef | xad | ggf |
|----------|-----|-----|-----|-----|-----|
| Distance | 78  | 113 | 21  | 94  | 27  |

From the above table it can be seen that there is no any relation between the distances (periods) of those strings, which means the distances between consecutive occurrences of the strings are not multiples and as a result the keyword is hard to be revealed.

#### V. CONCLUSION

In this paper, text encryption and decryption algorithm based on Hidden Encrypted Symmetric Key (HESK) is designed and implemented. The strength of this algorithm is represented in the key used for encryption and decryption process. The key itself is encrypted prior to be used for plain text encryption and cipher text decryption processes, then it is hidden inside the cipher text in such a way which makes it cannot be recovered. The aim of hiding the key is to overcome the problem of distributing the secret key and to make the proposed algorithms more secure and difficult or even impossible to be broken.

The proposed algorithm was tested on a set of plain texts of various sizes. The experimental result has been demonstrated that it is difficult to factorize the used key. The main two advantages of the proposed algorithm are represented in the computation simplicity and security efficiency.

#### REFERENCES

- [1] William Stalling, “Cryptography and Network Security Principal and Practice”, Third Edition, Pearson (2006).
- [2] Ayushi, “A Symmetric Key Cryptographic Algorithm”, International Journal of Computer Applications (0975 - 8887), Vol. 1, No. 15, (2010)
- [3] Arjen K. Lenstra and Eric R. Verheul., “Selecting cryptographic key sizes”. In Public Key Cryptography, pp 446-465. (2000).
- [4] Prashant Kumar Arya et al , “Comparative Study of Asymmetric Key Cryptographic Algorithms”, International Journal of Computer Science & Communication Networks, Vol 5(1),17-21 (2015)
- Deitel, H.M. and P.J. Deitel, “Java: How to Program”, Fifth Edition, Perentice Hall Inc., Upper Saddle River, New Jersey. (2002)

# A Distributed Security Situation Evaluation Model for Global Network

Weiwei Zhang

School of Cyber-Physical Systems and Control, Peter the Great St. Petersburg Polytechnic University, RUSSIA, St. Petersburg, 29 Polytechnicheskaya str., email: carlchueng@outlook.com

**Abstract:** Global network security assessment under distributed environment is extremely urgent. We try to design a distributed security situation quantitative evaluation model, and simulate the distributed security evaluation of the service subnet by building the LAN experimental platform. The result shows that this model has high practical value for vulnerability and attack means analysis of global network.

**Keywords:** network security, distributed security evaluation, centralized environment, appraisal procedure, global network

## I. INTRODUCTION

With the rapid growth of global trade, more and more economic investment has flowed into Central and Eastern Europe and the BRICS countries. Multinational corporations have set up branches or joint venture companies in these areas, which has also brought about certain cyber security risks while helping local governments increase the income and improve the employment rates. Large international groups and R&D institutions have a wide range of branch offices, as well as a complex network environment and ragged levels of security policies, so they tend to be targeted by hacking organizations. How to conduct a security assessment of a company's network from a global perspective is crucial for setting up an effective security strategy in the next step.

As J. McCumber said [1], the evolution of cybersecurity assessment method has gone from an artificial, local, single stage to an automated, global, and widely distributed situation. It is particularly noteworthy that the security of a host in a network depends not only on its own security status, but also the security of other ones in the global network. Therefore, assessing from the entire network is of great significance for discovering weaknesses in global network. F.B. Shaikh, and S. Haider [2], after analyzing the security threats of cloud computing, believe that the identification and analysis of distributed vulnerability are very important in the global environment, especially big data and cloud platform. Vulnerability is the direct premise of the security threat. No matter how advanced the attacker uses, if the protected assets have no weakness or only a slight vulnerability, it is difficult for the attackers to make use of their tools to damage assets [3,4,5]. Therefore, by identifying and analyzing the loopholes and security conditions of the services running on the network system, it is helpful to improve the level of network security protection and provide effective support measures for the integrated security management of the system.

However, the development of efficiency and security issues are often a pair of brothers who go hand in hand, and the whole

is usually not a simple sum of parts. With the rapid development of the Internet in today's information society, cross-domain collaboration and sharing between different branches of enterprises have become necessary tools for enterprises to improve their competitiveness and expansion [6]. Undoubtedly, this will expand the company's own virtualized network boundaries, making it easier for attackers to exploit various security vulnerabilities to implement distributed [7], springboard attacks, or to achieve successful intrusion through application layer trust relationships between security domains [8]. Therefore, the security problem of distributed systems is not a synthesis of the security problems of distributed nodes. Local intrusion detection and auditing are not enough to deal with the security threats brought about by the virtualization of the organizational structure [9]. The security policies between subsystems cannot be simply added. It is difficult to implement uniform security standards in different branches. From the perspective of the global network, different branches often combine into different subnets because of the need for information exchange. Subnets also need to interact with the outside world, such as the CRM system we are familiar with. Sales consultants need to be allowed to access the company's CRM system and get analytical support to conduct business with customer companies. Because this process takes place outside the company and lacks strong supervision, the service port of the system network can easily be abused or even hacked during this process. It can be said that the cross-domain collaboration and sharing between enterprises has raised higher and more specific requirements for enterprise security assessment and protection.

Therefore, it is necessary to analyze the problems of the existing network security assessment methods based on the actual needs of cross-domain situation, and design a distributed network security situation assessment model.

## II. THEORETICAL ANALYSIS

Under the actual demand of cross domain sharing, collaboration and defense of enterprise network, the existing network security assessment methods are faced with the following difficult problems:

1) The assessment of the security threat status of the network system usually focuses on the impact of the attack in a single network domain [10], which is difficult to reflect the global security threat situation, and is not conducive to the formulation and correction of the system security strategy.

2) When the enterprise has multiple different branches, in order to realize the security assessment of the business system from the whole point of view, the estimated business network

inevitably transfers private data to other sub network participating in the evaluation of the business, and this has become a privacy security problem [11].

Under the above background, based on the massive alarm information of IDS, the importance of service, the frequency of alarms, and the severity of security threats, we try to design a quantitative assessment model.

We use the centralized service security index  $F_{S_{jm}}$  to start the derivation of the model framework. Security index for service  $S_j$  in network  $m$  refers to the evaluation index of the losses caused by the intrusion using the vulnerable points.  $F_{S_{jm}}$  is derived from the importance of the service, the number of attacks  $C_{jtim}$  on the service  $S_j$ , and the severity of the attack  $P_{ji}$ . Based on the analysis method of literature [12], according to the characteristics of service operation in the network system, the importance of the service  $\theta_{jtm}$  is measured by the normal access of the system service in different time periods. Eq. (1) gives the calculation method for the service security index  $F_{S_{jm}}$  of the network  $m$ :

$$F_{S_{jm}} = \sum_{t=1}^h \theta_{jtm} \sum_{i=1}^{k_m} 10^{p_{ji}} C_{jtim} \quad (1)$$

More explanations for the formula:

### 1) Normal access $\theta_{jtm}$

The number of normal access  $\theta_{jtm}$  about service  $S_j$  varies from time to time during different time periods. Therefore, the same attack event has different influences and losses on services during different time periods. We can define the number of divided periods  $h=3$ , and divide the time of the day into three periods:  $\Delta_{t_1}=\text{Night}$ , which represents the time range of 0:00-8:00,  $\Delta_{t_2}=\text{OfficeHour}$  describes 8:00-18:00, and  $\Delta_{t_3}=\text{Evening}$  indicates the time interval from 18:00 to 24:00.  $\theta_{jtm}$  is assigned by the system administrator according to the normal average visit amount  $A_{jtm}$  ( $t \in [1 \dots h]$ ) of the service  $S_j$  in each period of the network  $m$ . The visit amount is represented by 1, 2, 3, 4, and 5 respectively: very low, low, medium, high, very high. The larger the value, the greater the average traffic. Then, we will obtain  $\theta_{jtm}$  in Eq. (2):

$$\theta_{jtm} = \frac{A_{jtm}}{\sum_{t=1}^h A_{jtm}} \quad (2)$$

### 2) Number of attacks $C_{jtim}$

We define the total number of types of services running in network  $m$  as  $d_m$ , count the number of alarms for different attack event types  $i$  ( $i \in [1, k_m]$ ,  $k_m$  is the total number of attack types for the corresponding service) of service  $S_j$  ( $j \in [1, d_m]$ ) according to the alarm data set generated by the IDS in the network. After generating the number of alarms, we can get  $C_{jtim}$ .

### 3) Security threat severity $P_{ji}$

After setting service  $S_j$  which suffers from different types of attacks  $i$  with severity  $P_{ji}$  during time period  $\Delta_t$ , we use the attack classification and prioritization of the SNORT user manual [13] to determine the threat severity of each attack. Respectively, 1, 2, and 3 indicate the three severity levels: low,

medium, and high. Table I is a partial attack category extracted from the SNORT user manual and its corresponding severity.

TABLE 1. ATTACK TYPE AND SEVERITY

| Attack category  | Description                                | Severity |
|------------------|--|----------|
| Attempted-admin  | Attempt to obtain administrator privileges | High     |
| Shellcode-detect | Executable code detected                   | High     |
| Successful-admin | Successfully acquired administrator rights | High     |
| Attempted-dos    | Attempt to cause a denial of service       | Medium   |
| Attempted-recon  | Attempt to cause Information disclosure    | Medium   |
| Network-scan     | Detected Network scan                      | Low      |
| String-detect    | Detected Suspicious string                 | Low      |
| Attempted-user   | Attempt to obtain User Rights              | High     |
| Trojan-activity  | Detected Internet Trojan                   | High     |
| Successful-user  | Successfully acquired User Rights          | High     |
| Misc-attack      | Mixed attack                               | Medium   |
| Suspicious-login | Suspicious user login                      | Medium   |
| Unknown          | Unknown traffic                            | Low      |
| Icmp-event       | General ICMP events                        | Low      |

## III. MODEL DESIGN

Expanding to a distributed environment, we assume that  $l$  ( $l \geq 3$ ) service subnets participate in the overall security assessment analysis, and there is no trusted third-party computing provider. Based on the historical alarm information collected by these subnets, the overall network service security index  $F_{S_j}$  can be calculated statistically. Because each evaluation participant has similar network services, by sharing the attack conditions of each service in its own network environment, under the distributed service assessment model, it obtains a more general and global security situation analysis result.

Suppose that the time division of the parties involved in the assessment is the same (day time is divided into three time periods: *Night*, *Office Hour*, and *Evening*), the same type of attack has the same security threat severity, and the total number of types of services running in the overall network is  $d$  ( $d \leq \sum_{m=1}^l d_m$ ). If the  $m$ -th party ( $m \in [1 \dots l]$ ) does not have an attack on a certain service type, the corresponding service  $C_{jtim} = 0$ . According to their respective IDS alarm

data sets, the parties count the  $F_{S_j}$  value of the entire network by sharing the service security index  $F_{S_{jm}}$  in their respective networks, so that we can get a globalized security posture result. The calculation method of the overall network service security index  $F_{S_j}$  is shown in Eq. (3):

$$F_{S_j} = \sum_{m=1}^l F_{S_{jm}} = \sum_{m=1}^l \left( \sum_{t=1}^h \theta_{jtm} \left( \sum_{i=1}^{k_m} 10^{P_{ji}} C_{jtim} \right) \right) \quad (3)$$

Based on the above methods, we present a quantitative assessment model (Fig.1) for distributed security posture.

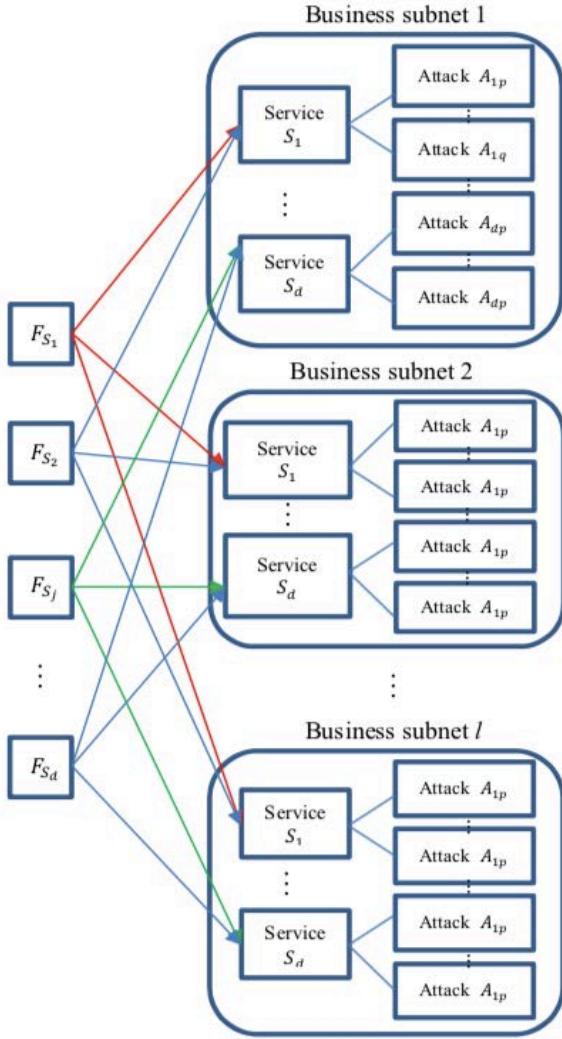


Fig.1. Distributed Security Situation Evaluation Model.

We can draw a conclusion that the greater the value of the service security index  $F_{S_j}$ , the higher the degree of security threat caused by exploiting the vulnerability of service  $S_j$ , which should be highly valued and prevented. Moreover,  $F_{S_j}$  also describes the security threat values for successive periods of time. The security threat trend of service  $S_j$  can be derived by comparing these values.

#### IV. MODEL CONSTRAINTS

In the process of computing the distributed security assessment, in order to statistics the multi-party analysis data,

the participants will inevitably transfer the private data to other participants to complete the distributed statistical process, resulting in privacy problems.  $F_{S_{jm}}$  describes an intrusion event that exploits the vulnerability of service  $S_j$  to attack a system. Therefore, network service information that is running or open in a network system is sensitive privacy information. The leakage of this kind of information may lead to the leakage and utilization of the system vulnerability information, which seriously affects the security of the service network.

At the same time, each business network participating in the evaluation needs to interact  $(l - 1) \times d$  times, when there are more participants or more types of services, the number of interactions will increase linearly.

#### V. MODEL VALIDATION

In order to verify the effectiveness of the proposed model in quantitative assessment of distributed security posture, we set up a LAN environment as an experimental platform to simulate the scenario of distributed comprehensive security assessment for three business subnets ( $l = 3$ ). Each subnet shares a class C address to connect to the Internet. Effective attacks on servers in each subnet are performed using intrusion methods such as buffer overflow and denial of service (DoS) attacks. In the experimental platform, the overall network service type number  $d = 5$ . SNORT is deployed on each server in the subnet, and the alarm information generated by it is used as the data source for security assessment. The network services running on the three subnets, as well as the distributed comprehensive service security index obtained from the statistics of one day's data, are shown in Table 2.

TABLE 2. SERVICE STATUS AND INTEGRATED SERVICE SECURITY INDEX(EXAMPLE)

| Net | Service | Service Importance<br>( $A_{t1}, A_{t2}, A_{t3}$ ) | Service Importance Weights<br>(3 periods)<br>$\theta_{t1}, \theta_{t2}, \theta_{t3}$ | Distributed Integrated Service Security Index |
|-----|---------|--|--|---|
| A   | FTP     | (1,4,3)  | (0.125,0.5,0.375)  | $F_{FTP} = 647.5$                             |
|     | MAIL    | (2,5,4)  | (0.182,0.455,0.364)  |   |
|     | DNS     | (1,3,2)  | (0.167,0.5,0.333)  |   |
|     | TELNET  | (1,2,1)  | (0.25,0.5,0.25)  |   |
| B   | MAIL    | (1,3,3)  | (0.143,0.429,0.429)  | $F_{DNS} = 337.6$                             |
|     | FTP     | (1,2,1)  | (0.25,0.5,0.25)  |   |
| C   | WWW     | (4,2,5)  | (0.364,0.182,0.455)  | $F_{TELNET} = 419.2$                          |
|     | TELNET  | (1,1,1)  | (0.333,0.333,0.333)  |   |
|     |         |  |  | $F_{WWW} = 863$                               |

The security threats of the five network services in the experimental platform within two weeks are shown in Fig.2. In our example, we can see that under the global conditions, the Internet, FTP, and MAIL are the more frequent attacks. Based on this, enterprise personnel can set corresponding policies, monitor FTP ports, set firewalls on external networks, and do a good job of Mail Anti-Phishing measures to improve the safety of the organization. It can be seen that the global

security threat posture map provides intuitive and quantitative data for the overall network security assessment. This method has high practical value for analyzing vulnerability, attack behavior and means of the whole network.

The parameters in the index are set in days. If the production system environment equipment is excellent, it is recommended that the enterprise security personnel set hours or even minutes.

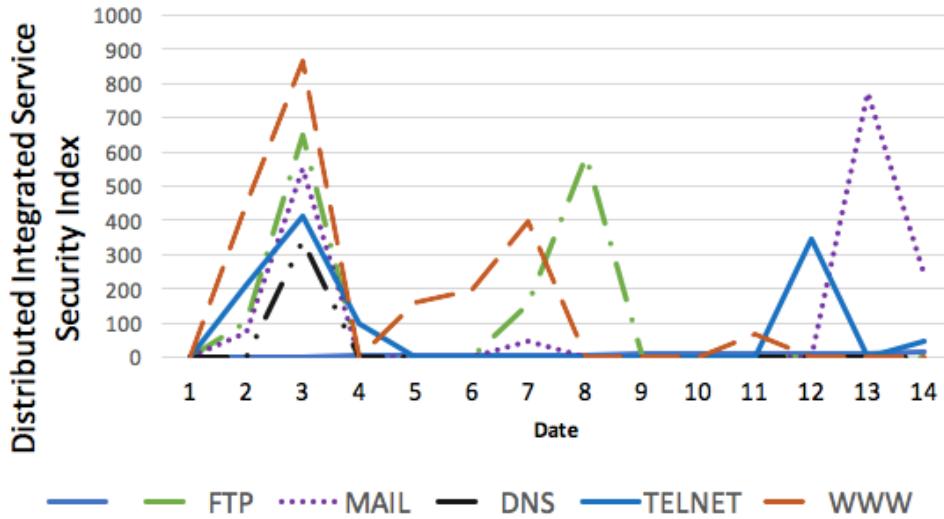


Fig.2. Global security threat Posture.

## VI. CONCLUSION

In order to solve the problem that it is difficult to use the massive and complex alarm information in the field of security assessment to effectively model the overall security situation, we combined the importance of services, the frequency of alarms, the level of security threats, and other factors, studied and proposed a distributed security quantitative assessment model, and verified the model at the same time based on massive alarm information. The results show that this model can perform distributed quantitative assessment under the condition of global security threats.

Secure distributed statistical model is a key issue for implementing network security assessment models in peer-to-peer environments. In the next research, we will try to analyze the privacy issues in distributed assessment, establish a distributed statistical model under the protection of privacy, and combine with different security assessment methods to support a wider range of application scenarios. Random number distribution and sane path methods are worth considering.

## REFERENCES

- [1] J. McCumber, " Assessing and Managing Security Risk in IT Systems: A Structured Methodology", *IEEE Trans. CRC Press*, pp.87-101, Apr. 2004.
- [2] F.B. Shaikh, and S. Haider, " Security threats in cloud computing", *ICITST*, 2011.
- [3] R. Raghavendra, K.B. Raja, S. Venkatesh, F.A. Cheikh and C. Busch, " On the vulnerability of extended Multispectral face recognition systems towards presentation attacks ", *ISBA*, 2017.
- [4] M. Almutairi and S. Riddle, "Security threat classification for outsourced IT Projects", *RCIS*, 2017, pp.447-448.
- [5] H.H. Wang, L.B. Shi and Y. Ni, "Distribution system planning incorporating distributed generation and cyber system vulnerability", *The Journal of Engineering*, pp.2189-2202, 2017.
- [6] K.W. Kongsgard, N.A. Nordbotten, F. Mancini, R. Haakseth and Paal E. Engelstad, "Data Leakage Prevention for Secure Cross-Domain Information Exchange", *IEEE Communication Magazine*, pp.37-43, 2017.
- [7] F. Hohl, "Automatically protecting computer system from attacks exploit security vulnerabilities", *Sony Corporation (Minato-ku, JP)*, 2007.
- [8] L. Zhou, Dan Wu, B. Zheng and M. Guizani, "Joint physical-application layer security for wireless multimedia delivery", *IEEE Communications Magazine*, Issue:3, pp.66-72, 2014.
- [9] Jiming Chen, Junkun and Ten H. Lai, "Energy-Efficient Intrusion Detection with a Barrier of Probabilistic Sensors: Global and Local", *IEEE Transaction on Wireless Communications*, Volume:12, Issue:9, pp.4742-4755, 2013.
- [10] M. Gharbaoui, F. Paolucci, A. Giorgetti, B. Martini and P. Castoldi, "Effective Statistical Detection of Smart Confidentiality Attacks in Multi-Domain Networks", *IEEE Transactions on Network and Service Management*, Volume:10, Issue:4, pp.383-397, 2013.
- [11] J. Giraldo, A. Cardenas and M. Kantarciooglu, "Security vs. privacy: How integrity attacks can be masked by the noise of differential privacy", *American Control Conference(ACC)*, 2017, pp.1679-1684.
- [12] X. Z. Chen, Q. H. Zhen, and X. H. Guan, "Research on Security Situation assessment of networked systems", *Journal of Xi'an Jiaotong University*, vol. 38, no. 04, pp. 404-408, 2004.
- [13] "SNORT", 2010; <http://www.snort.org/>.

# Information and Analytical Support of Economic Activity

# Using the Crowdsourcing Online-Platform as IT Tool for Gender Equality Plan Development

Ganna Plekhanova<sup>1</sup>, Olena Plokha<sup>2</sup>

Information Systems Department, S. Kuznets Kharkiv National University of Economics, UKRAINE, Kharkiv, 9-A Nauky avenue, email: [Ganna.Plekhanova@hneu.net](mailto:Ganna.Plekhanova@hneu.net)<sup>1</sup>, [Olena.Plokha@hneu.net](mailto:Olena.Plokha@hneu.net)<sup>2</sup>

**Abstract:** The paper is devoted to summarizing the experience of using the crowdsourcing online-platform as an IT tool to support the gender audit process and Gender Equality Plan development at universities on the example of Simon Kuznets Kharkiv National University of Economics (KhNUE). The paper considers the stages of the platform development, the strategy of attracting users, the results of the platform operation at KhNUE.

**Keywords:** crowdsourcing online-platform, gender audit, gender equality, Gender Equality Plan.

## I. INTRODUCTION

Gender equality policy today has become one of the most important trends in the development of the European higher education system. The issue of gender equality, that is, the parity of participation of both men and women in education, research, career development, university management and decision-making, is an important strategic task for modern European Universities, since ensuring gender equality in all spheres of social relations is one of the most important priorities of the development of the European Union (EU) [1], [2]. Ukraine, proclaiming a course on European integration, should make the best efforts to overcome the problem of gender discrimination and develop real mechanisms and tools for implementation of the gender equality strategy. Ukraine has ratified a number of international conventions on gender issues and has undertaken to guarantee equal rights and opportunities for women and men. That is why the issues of gender monitoring, analysis and audit are becoming more and more relevant for Ukrainian educational institutions. This led to the choice of the topic of this study, which aims to summarize the experience of using the online platform as an IT tool for the gender audit process support at higher education institutions in general and at IT departments, in particular.

## II. CROWDSOURCING PLATFORM AS A TOOL FOR SUPPORTING THE GENDER AUDIT PROCESS

The gender audit is a powerful and effective tool for forming a gender-sensitive environment at the university. The gender audit, based on the principle of active participation, is a modern monitoring tool for advancing gender equality policies, where not only quantitative but also qualitative indicators play a significant role. The peculiarity of such an audit is that self-assessment and democratic involvement is used, when all interested employees

voluntarily have the opportunity to speak on issues related to the gender perspective. The main purpose of the gender audit is to facilitate training, increase employee awareness and integrate gender perspectives into policy and current activities of the organization.

Consider the experience of conducting a gender audit of the university in general and the IT faculty in particular on the example of the Kharkov National Economic University named after Semen Kuznets.

At KhNUE gender audit was conducted with the financial and methodological support of the European Union (EU) as an integral part of the EQUAL-IST project "Gender Equality Plans for Information Sciences and Technology Research Institutions" [3] funded by the EU within the framework of the HORIZON2020 program.

The main objective of the three-year EQUAL-IST project is to implement structural changes to increase gender equality and positive changes in three main areas: first, in the processes of internal university management, including in personnel management processes, and secondly, in the processes of research design and delivery, thirdly, in the process of learning and interaction with students and in internal communications.

The methodology of gender audit used in the project was based on the International Labor Organization (ILO) methodology [4], but was substantially adapted to the needs of IT faculties of universities. It should be noted that the experience of adapting the ILO methodology to the needs of universities is not new for Ukraine, such work was carried out by the Network of Gender Centers of the universities of Ukraine, and the experience of adapting the methodology to the needs of IT departments is innovative not only for Ukraine, but also for Europe.

During the gender audit work was carried out in three main areas: firstly, personal interviews with university staff, and secondly, seminars with students, faculty members, representatives and heads of functional units (4 seminars in total). The third stage of work was related to the quantitative analysis of indicators of gender parity. In addition, a survey was conducted among participants of seminars. Main purpose of it was to identify gender inequalities and discrimination. All these measures allowed not only to identify problems with the implementation of equal opportunities for women and men, but also to identify the slots for their solution.

In order to receive more information from different sources in one place it was reasonable to use on-line

collaboration platform. As a collaborative workspace for gathering opinions of the different people from different countries within the Horizon 2020 EQUAL-IST project was suggested to develop and use crowdsourcing on-line platform CrowdEquality [5].

CrowdEquality is an interconnected environment in which all the participants in dispersed locations can access and interact with each other just as inside a single entity. The environment is supported by electronic communications and groupware, which enable participants to overcome space and time differentials.

The main goal of development of the crowdsourcing platform is to collect ideas and triggering discussions about promising initiatives for promoting gender equality and diversity and improving work-family balance in research institutions. CrowdEquality strives to be an inclusive community that empowers men and women all over the world to freely discuss the issues of gender equality in research institutions. CrowdEquality is designed to connect academic and non-academic staff members working at research institutions, policy makers, gender experts, members of relevant Non-Governmental Organisations (NGOs) and national/international networks, as well as all individuals interested in the design and implementation of Gender Equality Plans (GEPs).

### III. CATEGORIES OF USERS OF THE PLATFORM

The crowdsourcing online platform was used as an IT tool to support the gender audit process and ensure gender parity. This collaborative platform, developed and implemented by the EQUAL-IST project consortium, is the first European online platform that collects problems in achieving gender equality at universities, ideas for overcoming them and creating conditions for discussion. Platform users are able to:

- become acquainted with the problems of achieving gender equality existing at universities in European countries;
- express their own vision of gender inequality in various fields at universities;
- share thoughts and influence the internal policies of universities and research institutions in Europe.

When developing the platform, the following categories of users were implemented: internal users (representatives / representatives of the university community - those used to register the address in the university domain); external users (experts / experts on gender issues); guest (unregistered user); platform administrator; administrator from the university.

For internal users, the platform provides the following basic functionality:

- register on the platform;
- to place information on the existing problems in the field of gender equality at the university;
- express their ideas on overcoming gender equality issues; get acquainted with other proposed problems and ideas, comment on them;
- express your preferences for the most addictive ideas by pushing "Like"; vote for best ideas (available only at the voting stage).

Basic features for external users: register on the platform; comment on the problem; comment on ideas.

At the stage of designing the platform, the project team generated a list of functional requirements for the platform, which became the basis of the development process of the IT product. Since the platform is actually a tool for ensuring the development of gender equality plans, at different stages of the platform's operation, its functionality has changed: ideation phase - April 2017; reviewing phase - May 2017; voting phase - June 2017; development of gender equality policy.

At the initial stages of the platform it was possible to put problems and ideas to overcome them. Then at the "voting" stage these functions were blocked, instead, users had the opportunity to vote for the best ideas. At the stage of developing a gender equality policy (the final stage), users could use the accumulated data (problems, ideas, the number of votes for each idea), but all other platform features were not available.

After the platform was created and launched, the key task was to engage users (external and internal) to visit the platform in order to gather information about gender issues in a particular institution.

In order to involve external experts KhNUE team conducted the following activities:

presentation of the project and the platform at a meeting of the University's Academic Council;

presentation of the platform at workshops;

presentation of the platform at the meeting of the Information Systems Department and faculty meetings;

promotion through the personal contacts;

presentation of the platform for students at the meetings of student groups and during breaks between classes;

promotion through the internal mailing lists;

promotion through the university's web-resources (website and social media).

In order to involve external experts the following activities were conducted:

presentation of the platform at several scientific conferences;

presentation at the meeting of Kharkiv branch of All-Ukrainian Network of Gender Education Centers and promotion through the mailing list of this network;

presentation at the meeting of Kharkiv Platform "Culture of Equality";

presentation of the platform in the frame of TV interview about gender equality perspectives in Ukraine.

### IV. RESULTS OF THE USE OF THE PLATFORM AT KHNU

As a result of the gender audit and the use of the crowdsourcing online- platform, the main problems in the field of gender parity were identified and ranked as in the KhNUE and at other universities of the EQUAL-IST project consortium (Table 1). In addition, more than 30 ideas have been collected to overcome the problems found at KhNUE.

TABLE 1. RANKING OF PROBLEMS IN THE FIELD OF GENDER PARITY AT KHNU

| Challenges  | Rank of the challenge |
|---|-----------------------|
| “Work-life balance” problem   | 1                     |
| No opportunities for students and academic staff who became parents   | 2                     |
| “Glass ceiling” problem (men are mainly at the decision-making level, women - at the operational level)   | 3                     |
| Lack of resources – no budget for gender issues, no personnel, no committee responsible for gender equality   | 4                     |
| Lack of gender culture and tolerance  | 4                     |
| The gender issue is consequently not a part of any decision making process  | 5                     |
| No courses on gender issues   | 5                     |
| Absence of the Gender Education Center at the University  | 6                     |
| No gender expertise of teaching materials   | 7                     |
| Low share of girls at computer sciences and boys at economy sciences; Very low number of initiatives to attract girls to computer sciences and boys – to economy sciences | 7                     |
| The concentration of women in certain women's sectors of employment, primarily in the humanities and economic departments   | 8                     |
| Gender stereotypes and lack of awareness about gender equality issues   | 8                     |
| Gender insensitive communications; Gender-sensitive language adaptation is at the start phase   | 8                     |

At the stage of the gender equality policy development, the best ideas proposed on the platform were used by universities in developing gender equality plans. All ideas were divided into four categories according to the sections of the GEPs: HR and Management Processes; Research Design and Delivery; Teaching and Student Services; Institutional Communication.

As a result of the audit and the crowdsourcing online-platform it was found that, with a commitment to the principles of gender equality in general, the staff and students did not have a clear understanding of gender concepts and an integrated gender approach. Gender is still considered an additional, not a basic concept. It is necessary to make the gender approach "visual", "concrete", "clear" and "integrated".

In this sense, a particularly important is an information campaign aimed at the promotion of gender equality. The most often expressed opinion is that staff members and students do not feel the problem of gender inequality at the University or do not consider gender imbalance an important problem requiring special measures to address. Very popular was the opinion among users that the University has more

pressing problems that require urgent solutions as well as resources. Much more staff and students were concerned about “Work-life” balance problem, especially those who had young children and those who worked or studied at the second shift.

The gender audit and the use of the crowdsourcing platform allowed, among other things, to identify problems that were not related to gender in themselves, but caused problems in team relationships (for example, work-life balance problems), to change the perceptions of gender equality, to become aware of this subject matter In the future, this will create the effect of the chain reaction, when the acquired knowledge, representations are transmitted to those communicating with the carrier of such information.

In addition, gender audit, as an effective tool for improving socially-oriented policies, and the crowd-sourcing platform as a powerful and modern IT tool, will enhance the attractiveness of the university as an employer and as a place for learning, since advancing gender equality policies is, among other things, imaginative component.

Thus, the crowdsourcing platform and gender audit at the university proved to be an effective means for performing a number of important tasks: assessment, monitoring, communication, gender sensitive work methods, and education. Analysis of the ideas and comments placed on the crowdsourcing platform, the results of the electronic voting of the university communities allowed project team to develop the Gender Equality Plan for KhNUE for 2017 – 2019 (Table 2).

The actions envisaged by this plan are extended to two levels of intervention: University level or level of Information Systems Department.

TABLE 2. GENDER EQUALITY PLAN FOR KHNU

|  |   |
|--|---|
| Challenge 1: “Work-life balance” problem and lack of facilities for the balance support<br>Goal 1: Improve work-life balance of staff and students                         |   |
| Action 1   | Implement provisions for ensuring priority when drawing up the schedule of classes for academic staff having young children, children with disabilities, large families, for pregnant women |
| Action 2   | Carry out research on determining the need and resources for opening a children's room on campus  |
| Action 3   | Implement provisions for ensuring teleworking for academic staff having young children, children with disabilities, large families, for pregnant women                                      |
| Action 4   | Promote a culture of equal family responsibilities among students and staff   |
| Action 5   | Create and support section on the University website (and / or Intranet) with information on gender equality issues, work-life balance rights, provisions and regulations                   |
| Challenge 2: “Glass ceiling” problem (vertical segregation)<br>Goal 2: Encourage the achievement of gender equality (vertical principle) through information and awareness |   |

|  |  |
|--|--|
| raising activities   |  |
| Action 6   | Conduct seminars and round tables on gender equality for administrative staff  |
| Action 7.  | Make recruitment procedures and selection criteria transparent   |
| Challenge 3: Concentration of female staff in certain women's sectors of employment (horizontal segregation)                         |  |
| Goal 3: Encourage the achievement of gender equality (horizontal principle) through information and awareness raising activities     |  |
| Action 8   | Conduct information and awareness raising activities on gender equality issues for staff   |
| Challenge 4: Lack of gender equality machineries   |  |
| Goal 4: Ensure sustainability of GEP's actions by establishing gender equality machineries   |  |
| Action 9   | Establish a Commission on Gender Equality Issues and develop mechanisms for its functioning  |
| Challenge 4: Lack of gender equality machineries   |  |
| Goal 5: Ensure sustainability of GEP's actions by including gender equality goals and measures in the University strategic documents |  |
| Action 10  | Ensure the support of certain provisions of European Charter for Researchers by university staff   |
| Action 11  | Propose and implement changes to existing Collective Agreement of University Administration and Trade Union Committee in order to insure the implementation of GEP |
| Challenge 5: Gender issue is not in focus in decision-making process   |  |
| Goal 6: Make decision-making bodies informed and committed to the principles of gender equality                                      |  |
| Action 12  | Collect, analyse and publish gender disaggregated statistics of the University   |
| Challenge 6: Lack of gender culture and awareness about gender equality issues   |  |
| Goal 7: Raise awareness about gender equality issues, develop gender culture of academic staff and students                          |  |
| Action 13  | Conduct information and awareness raising activities on gender equality issues for students and academic staff   |
| Action 14  | Develop and disseminate printed and electronic awareness raising materials on gender equality  |
| Challenge 6: Lack of gender culture and awareness about gender equality issues   |  |
| Challenge 6: Lack of gender culture and awareness about gender equality issues   |  |
| Goal 8: Adopt a gender approach in specific teaching materials   |  |
| Action 15  | Perform pilot anti-discriminatory expert   |

|   |  |
|---|--|
|   | assessment of teaching materials   |
| Challenge 7: Gender imbalance among students of computer sciences and economy sciences      | Goal 9: Improve gender balance among students of computer and economy sciences   |
| Action 16   | Conduct awareness raising events to spread information about women in IT industry and women's career opportunities in IT           |
| Action 17   | Establish special nominations for women / female teams in existing IT championships and competitions                               |
| Action 18   | Perform pilot anti-discriminatory expert assessment of marketing and advertising materials for IT Bachelor's and Master's programs |
| Challenge 8: Gender insensitive communications  |  |
| Goal 10: Raise awareness about the value of gender sensitive / gender neutral communication |  |
| Action 19   | Provide language and visual support for gender equality in the media content of the University                                     |

### III. CONCLUSION

The further work of the EQUAL-IST project team will be related to the implementation of GEP and the dissemination of experience among other universities. The issues and ideas proposed on the platform will continue to serve as a basis for developing gender equality plans for universities and for disseminating the best practices of European higher education institutions.

### ACKNOWLEDGMENT

This research was conducted in the frame of the EQUAL-IST project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 710549. Unit EU.5. – Science with and for society.

### REFERENCES

- [1] Strategic engagement for gender equality 2016-2019, [http://ec.europa.eu/justice/gender-equality/document/files/strategic\\_engagement\\_en.pdf](http://ec.europa.eu/justice/gender-equality/document/files/strategic_engagement_en.pdf)
- [2] Strategy for equality between women and men 2010-2015, [http://ec.europa.eu/justice/gender-equality/files/documents/strategy\\_equality\\_women\\_men\\_en.pdf](http://ec.europa.eu/justice/gender-equality/files/documents/strategy_equality_women_men_en.pdf).
- [3] EQUAL-IST project, <https://equal-ist.eu/overview>
- [4] A manual for gender audit facilitators: The ILO participatory gender audit methodology, 2nd Edition, [http://www.ilo.org/wcmsp5/groups/public/-/-dreports/-/-gender/documents/publication/wcms\\_187411.pdf](http://www.ilo.org/wcmsp5/groups/public/-/-dreports/-/-gender/documents/publication/wcms_187411.pdf)
- [5] CrowdEquality - the Idea Crowdsourcing Platform for Promoting Gender Equality and Diversity, [www.crowdequality.eu](http://www.crowdequality.eu)

# Advances in GMDH-based Predictive Analytics Tools for Business Intelligence Systems

Serhiy Yefimenko

Department for Information Technologies of Inductive Modelling, International Research and Training Center for Information Technologies and Systems, UKRAINE, Kyiv, 40, Ave Glushkov, email: syefim@ukr.net

**Abstract:** The paper analyzes approaches to prediction of economic processes in business intelligence systems. Contemporary tools of predictive analytics, used for effective making of business decisions, are considered. The concept of advanced GMDH-based predictive analytics tool is proposed.

**Keywords:** business intelligence, predictive analytics, GMDH, recurrent-and-parallel computing.

## I. INTRODUCTION

Achieving success and ensuring competitiveness in today's fast changing economic conditions are impossible without the use of reliable and on-line information. Business data is becoming significant resource for knowledge acquisition and making important managerial decisions in different business fields. Up-to-date effective decisions require reliable and complete information, and it is impossible to do with the use of traditional information systems.

In our time, there is a rapid transformation of the global information area that affects society, market and business. There is a fast growth of the digital economy. 25% of the world's economy will be digital by 2020 [1], whereas this number was 15% in 2005. The Internet of Things (IOT) and Big Data, mobile and cloudy technologies contribute the economy digitization. Influence of these technologies on business will result in direct domain physical resources to become useless.

Business intelligence (BI) is a modern managerial tool in the digital economy. It contributes to the company's prosperity based on smart financial, business processes, and personnel management under considerable amount of information.

The purpose of the review is to consider modern approaches to prediction economic, production and financial processes in BI systems, as well as existing software tools for predictive analytics.

## II. PREDICTIVE ANALYTICS & PREDICTIVE MODELLING

BI encompasses strategies and technologies used by enterprises to analyze business information [2]. BI refers to the management philosophy and toolkit used to help operate business information in order to make effective business decisions. BI technologies provide historical, current and predictive views on business operations.

The classification of technologies, used by business analytics, is given in [3]. Predictive modeling is one of the most effective.

Organizations of different types may be troubled by certain problems in the effectiveness of existing data using in their systems. In this regard, the quality and speed of information and analytical support for business management is of particular importance for companies. Most of them use BI analytical applications based on OLAP systems for planning, analyzing and controlling tasks. However, in new economic conditions, the functionality of such systems is not enough to solve new digital problems, since they oriented on retrospective analysis. Consequently, there is a need for predictive analytics, which complements and enhances BI capabilities in terms of predicting future events.

In general, there are several types of analytics that co-exist and supplement each other [4, 5]:

– *descriptive analytics* explores past facts in order to find the causes of previous successes or failures. It answers the question "What's up?". Descriptive analytics is still in use today. Most of the management reports for sales, marketing, finance use this kind of business analytics;

– *diagnostic analytics* goes further and gives an idea not only of the events that occurred, but also of their causes. It answers the question "Why something happened?";

– *predictive analytics* answers the question "What is likely to happen?". Historical data is combined with rules, algorithms and external data in order to determine the future value or the probability of an event;

– *prescriptive analytics* is the next stage in predicting future events, and offers a sequence of actions to gain most from predictions and shows the consequence of each decision. It answers the question "What should I do?".

Predictive analytics is defined in [6] as a variety of statistical techniques from predictive modelling, machine learning, and data mining that analyze current and historical facts to make predictions about future or otherwise unknown events. As a rule, big data arrays are used in the process of analysis. The main idea of predictive analytics is to determine one or more parameters that affect the predicted event. The process of predictive analysis can be represented as follows:

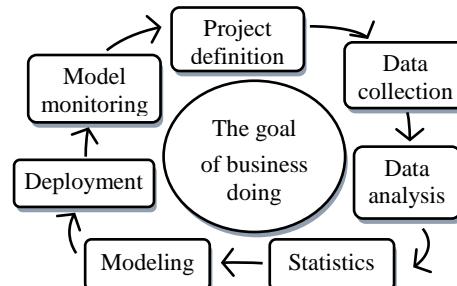


Fig.1. Predictive analytics process

*Project definition.* Definition of project results, components, scale of the work, business purpose, data set to be used.

*Data collection.* With the use of intelligent data analysis, data from different sources is prepared.

*Data analysis.* The process of data checking, clearing and modeling in order to identify useful information is performing.

*Statistical analysis* allows to confirm assumptions, hypotheses using standard statistical models.

*Predictive modeling* provides the ability to automatically build accurate predictive models.

*Deployment* of a predictive model provides the use of analytical results in the decision making process for obtaining reports.

*Model monitoring.* Models are tested to ensure the expected results.

The result of predictive analytics applying consists in the most effective business solutions making. An important requirement for a predictive model is to be as fit as possible and to be statistically significant. The predictive models may be [7]:

- *classification models*. They describe set of rules, according to which a new object can be assigned to the relevant class;

- *time series models*. They describe the functions that allow prediction of continuous numerical parameters and are based on information on the change of a certain parameter over the past time period.

According to Transparency Market Research [8], the market for predictive analytics will reach \$ 6.5 billion by 2019, while it was \$ 3.6 billion in 2015. The global market for predictive analysis systems will grow by an average of 17.8% annually. And as experience shows, the companies survive, that continue to invest in technology and innovation in the difficult economic times. And predictive analytics, of course, is one of such technology.

### III. SOFTWARE TOOLS FOR PREDICTIVE ANALYTICS

Forrester Research has published in 2013 a report "Big Data Predictive Analytics Solutions, Q1 2013" in which market leaders for predictive analytics are contained [9]. According to it, SAS and IBM SPSS have the strongest position in the market and the best strategies among the largest developers of predictive analytics tools. The evaluation was carried out for 51 parameters - from the completeness of the functionality for the main analytical system to the size of the client base and the architectural advantages offered by the solutions developers.

SAS (Statistical Analysis System) *Enterprise Miner* [10] is leading in the segment of in-depth analytics, accounting for about a third of the market. It allows users to explore and analyze large amounts of data, to find patterns of relationships and to make well-informed decisions, based on facts and findings. Areas of effective use of the solution: banking sector, healthcare, oil and gas sector, insurance companies, telecommunications, transport, power system.

The main advantages of SAS Enterprise Miner include:

- advanced predictive modeling;

- convenient and clear interface allowing users to create predictive models on their own;
- automated process of routine application of models;
- possibility of batch processing;
- rapid data collection and preparation, aggregation and analysis;
- scalability and customization of the solution;
- high system performance when working with big data.

*IBM SPSS* (Statistical Package for the Social Sciences) [11] is a widespread intelligent tool for predictive analytics. SPSS's predictive analytics helps you analyze the patterns in historical and current transactions to predict potential future events.

A key component of the toolkit is *SPSS Modeler*, software environment for data mind allowing you to create intelligent predictive solutions by revealing the data patterns and relationships. *SPSS Modeler Server* supports integration with data mind and modelling tools provided by DBMS (database management system) developers, including IBM Pure Data System for Analytics. Using the SPSS Modeler, one can build and store models in the database. One can combine the analytical capabilities and ease of use of SPSS Modeler with the power and performance of the DBMS, using the built-in algorithms supplied by their developers. The models are built inside databases and are available for use with the convenient user interface of SPSS Modeler.

*Dell Statistica* (from 2017 -Tibco Software) in-depth data analysis platform [12] focuses on data professionals and organization needing to data process from a large number of IOT devices and heterogeneous sources. The functionality of the toolkit will help to prepare structured and unstructured data, deploy analytical tools on devices regardless of their location and use internal analysis functions on the MYSQL, Oracle, and Teradata platforms.

With Dell Statistica, companies are able to cope with the lack of data analysts and the complexity of today's IOT environments, as well as take into account new sources and data types.

*Dell Statistica*'s features, simplifying predictive analytics, are as follows:

- dashboards with advanced visualization allowing users to easily see the results of the analysis at any stage;
- state-of-the-art web interface allowing users to share reports that can be opened in any browser;
- effective control of data, entered manually.

In addition to the represented (far from complete) developers of predictive analytics, there are also a large number of specialized firms providing business intelligence services. One of the most famous is Elder Research [13]. It has extensive experience in using many software tools (including all the above) for developing analytical solutions, programming, and personalized data visualization.

### IV. GMDH-BASED PREDICTIVE ANALYTICS TOOLS

Among the various tools for predictive analytics, it should be emphasized several ones, the common feature of which is using of one of the most effective inductive modeling methods – Group Method of Data Handling (GMDH) [14].

Software tool *Insights* [15] is developed by German company *KnowledgeMiner Software* (created in 1993). Besides GMDH, it also uses Similar Patterns self-organizing modeling technology (also known as Analog Complexing) and fuzzy logic for modelling and prediction. It is possible to build linear and nonlinear, static and dynamic time series models, multi-input and one output models, many inputs and many outputs models. The outputs of the model can be represented both in analytical form (in the form of equations with estimated parameters) and graphically (using a system graph, which reflects the interconnections of the system structure).

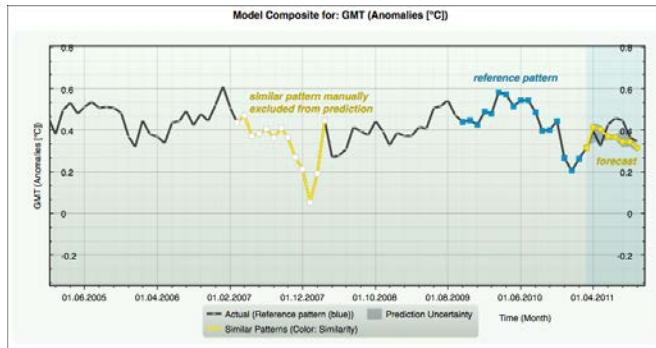


Fig.2. An example of Insights using for predictive analysis

Insights implements vector processing, multi-core and multiprocessor support for high-performance computing. It is scaled to the Apple Macintosh computer hardware. Regardless of which processor is used (dual-core or two six-core), the software automatically uses all the features of the PC.

*GMDH Shell* [16] is a contemporary software tool for predictive analytics. It is based on the classical GMDH algorithm and can be used for time series prediction, solving classification and clustering problems. *GMDH Shell* is a powerful solution for analyzing multidimensional data from various business fields. The software tool offers data mining algorithms – self-organized neural networks and combinatorial structural optimization of models. There is also the possibility of high-performance computing using a Linux-cluster.

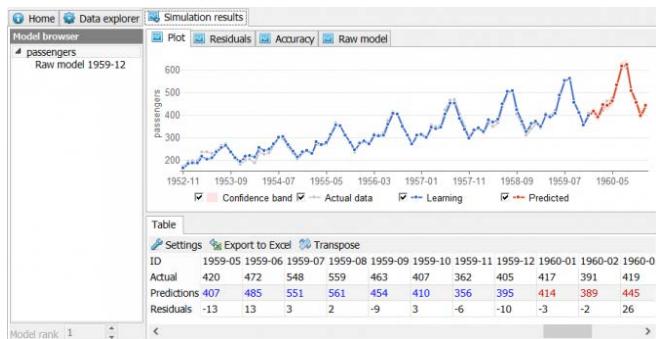


Fig.3. An example of GMDH Shell using for predictive analysis

It should be noted that GMDH Shell does not compete with *KnowledgeMiner Insights* in the sense that it is intended for use on the Windows operating system.

*Software tool* for modeling and prediction of complex multidimensional interrelated processes is developed in the

Department for information technologies of inductive modelling [17].

The tool is implemented for use on multiprocessor cluster systems. However, it can be embedded in any contemporary business intelligence system as an analytical tool for modeling and prediction of the dynamics processes in digital economy systems based on the detection and use of knowledge about the behavior and performance of such systems.

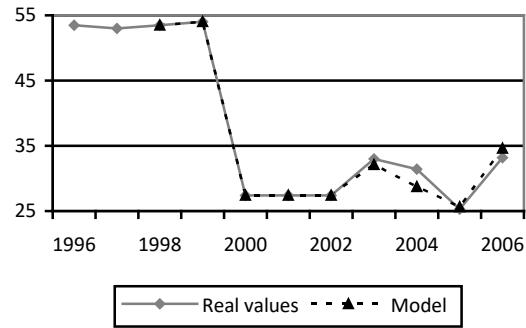


Fig.4. An example of software tool with recurrent-and-parallel GMDH algorithms using for predictive analysis

## V. ON CONSTRUCTION OF ADVANCED GMDH-BASED PREDICTIVE ANALYTICS TOOL

Given that there is a considerable amount of predictive analytics tools, not fulfilling the whole range of problems, it may be concluded that there is still no single convenient solution on the market. To create accurate models and to obtain adequate predictions of various indicators, an advanced predictive analytics tool is required allowing to comprehensively reflect the relationships in models, being accessible and convenient for the user, allowing the user to customize the model and build reliable predictions. However, to date there is no such a complete solution. Developing of such a tool is an actual problem in the field of analytical business solutions.

The most important features of this advanced predictive analytics tool are:

- GMDH-based software tool [18, 19];
- recurrent-and-parallel computing [20];
- intelligent user interface [21];

*GMDH-based software tool*. The user does not need to have a thorough knowledge of the modeling principles when building models. He will be able to model with a convenient tool, knowing only the features of his domain. Built-in intelligent algorithms allow one to automatically build models on the available data set, which greatly facilitates the user's work.

Advanced predictive analytics tool, being constructed in the paper, is based on software for modelling and prediction of complex multidimensional interrelated processes in the class of vector autoregressive models.

*Recurrent-and-parallel computing*. Fundamentally new data-based solution for inductive modeling of complex processes has a high level of performance because of new concept, combining the efficiency of recurrent and parallel computing. The implementation of such solution provides

significant enhancing of efficiency and validity of managerial decisions.

*Intelligent user interface.* It is very important that predictive analytics tools are either too complicated for users or do not contain the necessary range of options. The intelligent user interface should be friendly and should allow building models without deep programming knowledge, which will significantly expand the range of users and increase their confidence in BI applications.

Advanced predictive analytics tool must include an intelligent shell allowing user (with any level of qualification) help to solve the data-based modelling problem (from data preprocessing to modelling algorithm choice). The intelligent shell provides the general use of automatic analysis and modeling procedures. It takes into account the user's wishes and a priori knowledge about the modeling object, and also provides decisions making control at every step of the modeling problem solving.

## VI. CONCLUSION

Contemporary capabilities and advanced techniques of predictive analytics are becoming powerful way for increasing the company's productive efficiency. Predictive analytics is a new trend opening up broad prospects for the further development of companies.

Applying predictive analytics systems one should understand that the work of such systems is impossible without sufficient historical data and ineffective without the collection of current data. The less data will be used, the less accurate are predicted values.

The effectiveness of applying predictive analytics tools depends on both technologies used and the quality of such tools. And the advantage here will be on the side of the solutions, providing advanced methods of data mining. Such ones are just knowledge-oriented intelligent modeling software tools based on GMDH.

## REFERENCES

- [1] [https://www.accenture.com/t20160314T114937\\_\\_w\\_\\_/us-en/\\_acnmedia/Accenture/Omobono/TechnologyVision/pdf/Technology-Trends-Technology-Vision-2016.PDF.P](https://www.accenture.com/t20160314T114937__w__/us-en/_acnmedia/Accenture/Omobono/TechnologyVision/pdf/Technology-Trends-Technology-Vision-2016.PDF.P).  
B. Johns, "A symmetrical condensed node for the TLM method," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-35, pp.370-377, Apr. 1997.
- [2] [https://en.wikipedia.org/wiki/Business\\_intelligence](https://en.wikipedia.org/wiki/Business_intelligence).
- [3] M. Goebel and L. Gruenwald, "A survey of data mining and knowledge discovery software tools", Volume 1, Issue 1 (June 1999), Publisher ACM New York, NY, USA.
- [4] [https://en.wikipedia.org/wiki/Business\\_analytics](https://en.wikipedia.org/wiki/Business_analytics).
- [5] <https://www.gartner.com/it-glossary/diagnostic-analytics>.
- [6] [https://en.wikipedia.org/wiki/Predictive\\_analytics](https://en.wikipedia.org/wiki/Predictive_analytics).
- [7] <http://www.globalcio.ru/workshops/968>.
- [8] <https://www.transparencymarketresearch.com/pressrelease/predictive-analytics-industry.htm>.
- [9] <https://www.forrester.com/report/The+Forrester+Wave+Big+Data+Predictive+Analytics+Solutions+Q1+2013/-/ERES85601>.
- [10] [https://www.sas.com/ru\\_ua/software/enterprise-miner.html](https://www.sas.com/ru_ua/software/enterprise-miner.html).
- [11] <https://www.ibm.com/analytics/data-science/predictive-analytics/spss-statistical-software>.
- [12] <https://www.tibco.com/products/tibco-statistica>.
- [13] <https://www.elderresearch.com>.
- [14] Stepashko V. "Developments and Prospects of GMDH-Based Inductive Modeling" In: *Advances in Intelligent Systems and Computing II: Selected Papers from the International Conference on Computer Science and Information Technologies, CSIT 2017*, September 5-8, Lviv, Ukraine. N. Shakhovska, V. Stepashko Editors. AISC book series, Volume 689. – Cham: Springer, 2017, pp. 474-491.
- [15] <https://www.knowledgeminer.eu>.
- [16] <https://gmdhsoftware.com>.
- [17] <http://www.mgua.irtc.org.ua>.
- [18] Yefimenko S. "Building Vector Autoregressive Models Using COMBI GMDH with Recurrent-and-Parallel Computations" In: *Advances in Intelligent Systems and Computing II: Selected Papers from the International Conference on Computer Science and Information Technologies, CSIT 2017*, September 5-8, Lviv, Ukraine". N. Shakhovska, V. Stepashko Editors. AISC book series, Volume 689, Cham: Springer, 2017, pp. 601-613.
- [19] Stepashko V.S. and Yefimenko S.M. "Technologies of Numerical Investigation and Applying of Data-Based Modeling Methods" *Proceedings of the 2nd International Conference on Inductive Modelling ICM 2008*, Kyiv, 2008, pp. 236-240.
- [20] Serhiy Yefimenko, Volodymyr Stepashko "Intelligent Recurrent-and-Parallel Computing for Solving Inductive Modeling Problems" *Proceedings of 16th International Conference on Computational Problems of Electrical Engineering (CPEE-2015)*, September 2-5, 2015, Lviv, Ukraine, 2015, pp. 236-238.
- [21] Stepashko V.S., Zvorygina T.F., Yefimenko S.M. "Problem of decision making intellectualization in tasks of models identification" (Problema intelektualizatsii pryiniatiia rishen' u zadachakh identyfikatsii modelei), *Proceedings of ISDMIT-2005 Conference*, Kherson, Ukraine, 2005, Vol. 1, pp. 127-131.

# Management Software Implementing within Ukrainian SMEs: Case Study

Vasyl Hryhorkiv<sup>1</sup>, Lesia Buiak<sup>2</sup>, Andrii Verstiak<sup>3</sup>, Mariia Hryhorkiv<sup>4</sup>, Olexandr Savko<sup>5</sup>

1. Department of Economic modelling and business informatics, Chernivtsi National University, UKRAINE, Chernivtsi, 2 Kotsyubynsky str., email: [v.hryhorkiv@chnu.edu.ua](mailto:v.hryhorkiv@chnu.edu.ua)
2. Department of Economic Cybernetics and Informatics, Ternopil National Economic University, UKRAINE, Ternopil, 3 Peremohy sq., email: [lesyabuyak@ukr.net](mailto:lesyabuyak@ukr.net)
3. Department of Economic modelling and business informatics, Chernivtsi National University, UKRAINE, Chernivtsi, 2 Kotsyubynsky str., email: [a.verstyak@chnu.edu.ua](mailto:a.verstyak@chnu.edu.ua)
4. Department of Economic modelling and business informatics, Chernivtsi National University, UKRAINE, Chernivtsi, 2 Kotsyubynsky str., email: [m.hryhorkiv@chnu.edu.ua](mailto:m.hryhorkiv@chnu.edu.ua)
5. Department of Economic modelling and business informatics, Chernivtsi National University, UKRAINE, Chernivtsi, 2 Kotsyubynsky str., email: [o.savko@chnu.edu.ua](mailto:o.savko@chnu.edu.ua)

**Abstract:** By this research we investigate implementation of management software systems within Ukrainian SMEs. We have pointed that implementing of ERP systems in such cases have many disadvantages according to economic crisis in Ukraine. By this case research, we propose alternative approach - the implementation of so-called scalable information systems-designers. Good examples of such systems are Accent 2 and 7.4. These systems are developed by Ukrainian IT-companies and enable complex automation of enterprises of any kind of activity.

**Keywords:** ERP, manufacturing enterprises, Management Software, Accent 2.

## I. INTRODUCTION

As thing stand, SME's (small and medium enterprises) can't go through in a modern competitive environment without electronic gadgets, remote access to corporate resources, automated planning and process management,. This is an obvious trend: companies of different profiles, regardless of the size of the business, switch to automated computer systems that help to conduct business, monitor and plan it. Efficiency from this noticeably wins, especially in terms of saving time and money.

The basis of automaton in commercial organizations is ERP systems (Enterprise resource planning).

ERP systems are implemented with the purpose of uniting all the sectors of the company and all necessary functions in one management system which is to meet the requirements of all these sectors. The development of such a system is not an easy task. ERP possesses the uniform data base of all sectors and tasks which means that the access to the information is much simpler, the basic advantage being the mutual information exchange among the sectors.

The implementation of the management system in many cases enables not so much to increase the profitability as to decrease the expenses. The managers who take the decisions have comprehensive information, may interpret it correctly and undertake the right actions. The financial profit lies very often in the fact that the authorized persons can manage more

effectively the production stock, decrease its quantity to meet the demands and, thus, release the circulating assets. The automation saves more time for the decision-making employees for analytical work as many time-consuming routine processes are being cut down. Besides, the enterprise with the automated management system meeting the requirements of the Western standards has good chances for the Western investments.

There are many researches focusing on advantages of ERP, listed in [1, 2, 3]. Notwithstanding, as stated in [2], despite the great recognition and acceptance of ERP Systems in organizations, some criticisms have been directed to these types of systems [4], pointing inflexibility [5-8] and long implementation period [5, 6, 9] and high cost.

Implementing of ERP systems by Ukrainian enterprises has special value as the tasks of the economic growth, increasing of production quality and competitiveness, becoming a member of the world market are most topical. The solution of these tasks without the increase of the quality management and automation of all spheres of economy (state bodies of management, different enterprises and small businesses) is impossible. But stated above disadvantages of ERP-systems including its high cost and primaries of SME's in Ukraine demand finding other solution.

In this paper we describe alternative approach of Enterprise management system implementation – using so-called scalable information systems-designers at the case of enterprise of wood processing industry.

## II. MARKET TRENDS

According to [10] there are top 10 ERP solutions for Ukrainian enterprises: OneBox, 1C: ERP, MS Dynamics ERP, IT-Enterprise, Parus, BAS ERP, DeloPro, HansaWorld, Galaktika, Tend Erp.

The market of ERP-systems is the part of the information and communication industry, which, in turn, brings together the two main sectors: Communications and Information Technology. As we stated in [11] one important factor of increasing ERP system demand is the currently ongoing

process of consolidation inside of several economic branches. Newly groups of companies and holdings need sufficient software and licenses in order to allow them to integrate different information systems into one management control system. According to estimations of the Ukrainian IT consulting enterprise IDC a high potential of growth have such ERP systems in industrial sectors within sales and distribution, banking sector and governmental and public administration. The speed of growth of these sectors will be higher than Ukrainian ERP market growth in general.

Well known investigation of Panorama Consulting and its report "Clash of the Titans 2017" sates that Microsoft Dynamics has passed by Oracle to reach the second-from-the-top spot in the ERP market share rankings published by Panorama Consulting [12]. According to Panorama consulting [13] SAP remains leading at the global market with a share of 19% in implementation of management software solutions, it is followed by solutions from Microsoft Dynamics (16% of the market) and Oracle (13%). There are also popular ERP solutions from Epicor and Infor.

We want to stress that there are no up-to-date data concerning Ukrainian ERP market, but the leaders in 2016, according to IDC research [14], were solutions from SAP (43.4%), IT-Enterprise (15.7%), 1C (13.9%), Oracle (11.7%) and Microsoft Dynamics NAV (6.1%). In recent years, according to a number of experts' investigations, the impact of 1C solutions has significantly decreased at the Ukrainian market as it is under sanctions [15-16].

In the above section we underlined disadvantages of ERP-solutions implementation by Ukrainian SME's accordingly to protracted economic crisis and its high cost. As alternative for this enterprises in Ukraine there are mane solutions based on so-called scalable information systems-designers. Good examples of such systems are Accent 2 and 7.4. These systems are developed by Ukrainian IT-companies and enable complex automation of enterprises of any kind of activity. A typical set of functional requirements provides management and financial accounting, but open source framework allows coding modules for complex automation. Implementation of A2 and Accent systems is achievable in stages, progressively covering all areas of a company [17-18].

In this paper we pay our attention to automation of manufacturing enterprises as it is a complex multistage process and the management information system of suchlike enterprise should include a number of modules majoring in certain areas. It is clear that comprehensive and complex automation of production could be provided through implementation of ERP systems.

### III. RESULTS

An example of mentioned problem is automation of company providing furniture manufacture. Information flows of the company can be divided into several interconnected modules. Implementation of the system takes place in parallel in all areas of their subsequent integration. Automation modules can be divided into several key blocks in simplified form (Fig. 1).

Each of the blocks has both incoming and outgoing information. The first step of automating is the accounting of

raw materials. The production process starts with the reception of raw materials that is one of the most important components of the manufacturing process. Typically, supplier ships raw materials providing the appropriate enclosed document with a complete list of materials specifying size and its volume.

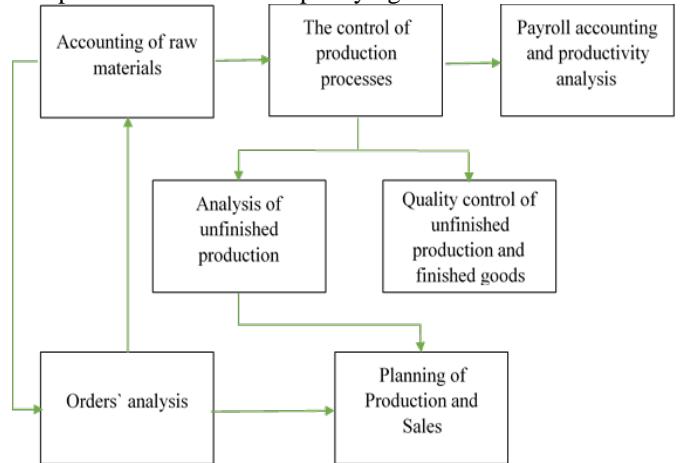


Fig. 1. Modules of information system

Responsible person performs the necessary measurements and marking of the timber party. This data is moving to the production department, which carries out a comparison of the information given in purchase order actual numbers. The next stage of the production process is the primary processing of wood. According to the established production plan, senior master production manager distributes raw materials as needed.

Primary wood processing involves a number of interrelated processes such as drying (if the raw materials come in raw form), longitudinal and transverse cutting (Fig. 2 and Fig. 3).

| Long grain №9 |                |       |              |        |          |              |       |
|---------------|----------------|-------|--------------|--------|----------|--------------|-------|
|               |                |       |              | 1      | date     | 01 June 2017 |       |
| From          | Cold sawmill   |       |              |        |          |              |       |
| To            | Cold sawmill   |       |              |        |          |              |       |
| Nº            | Bunch Nº       | Name  | Wood species | Unit   | Q-ty, m³ | Amount       | Size  |
| 1 389         | Board beech 57 | Beech | m³           | 1,79   | 1.7900   | 3000         | 57 25 |
| 2 400         | Board beech 57 | Beech | m³           | 2      | 2.0000   | 3000         | 57 25 |
| 3 396         | Board beech 57 | Beech | m³           | 1,93   | 1.9300   | 3000         | 57 25 |
|               |                |       |              | 0      | 0        | 0            | 0 0 0 |
| Total         |                |       |              | 5.7200 | Total    | 5.7200       | Total |
|               |                |       |              |        |          | Beech        | 5.72  |
|               |                |       |              |        |          | Oak          | 0     |

Fig. 2. Longitudinal board cutting

The entire production process is displayed in the system by means of relevant documents that allows operational control of the production process. The final stage of primary wood processing is the formation of rough pieces and transfer them to areas of further processing.

02. Cross grain of board - A2

**Cross grain №** [ ] date [ ] 01 June 2017 [ ]

From [ ] Cold sawmill  
To [ ] Cold sawmill

| Wood species | Size                  |                |                |      |     |    | Beech №      | 0          |           |          |        |
|--------------|-----------------------|----------------|----------------|------|-----|----|--------------|------------|-----------|----------|--------|
|              | L                     | W              | T              |      |     |    |              |            |           |          |        |
| Beech        | 3000.00               | 100.00         | 32.00          |      |     |    | Oak №        | 7          |           |          |        |
| Nº           | Name                  | Unit base      | Unit der.      | Size |     |    | Wood species | Series     | Q-ty, pcs | Q-ty, м3 |        |
|              |                       |                |                | L    | W   | T  |              |            |           |          |        |
| 1            | RSS Beech 1850x100x32 | шт             | м <sup>3</sup> | 1850 | 100 | 32 | Beech        |            | 242.000   | 1.4326   |        |
| 2            | RSS Beech 450x100x32  | шт             | м <sup>3</sup> | 450  | 100 | 32 | Beech        |            | 979.000   | 1.4098   |        |
| 3            | RSS Beech 800x100x32  | шт             | м <sup>3</sup> | 800  | 100 | 32 | Beech        |            | 482.000   | 1.2339   |        |
| 4            | RSS Beech 100x32      | м <sup>3</sup> | м <sup>3</sup> |      | 100 | 32 | Beech        | Ч(01/06)Б2 | 1.1800    | 1.1800   |        |
| 5            | RSS Beech 50x32       | м <sup>3</sup> | м <sup>3</sup> |      | 50  | 32 | Beech        | Ч(01/06)Б3 | 0.8400    | 0.8400   |        |
|              |                       |                |                |      |     |    |              |            |           | Total    | 6,0963 |
|              |                       |                |                |      |     |    |              |            |           | Beech    | 6,0963 |
|              |                       |                |                |      |     |    |              |            |           | Oak      | 0      |

[ ] cross grain document / < > CAPS NUM

Fig. 3. Transverse board cutting

The conversion process of transforming of rough-sawn stock into finished products is shown in Fig. 4.

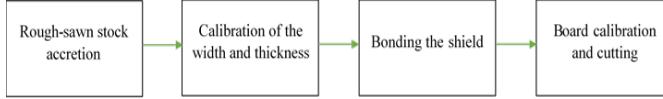


Fig. 4. Assembling of the finished product

Rough-sawn stock hits the seam line after initial treatment, where they to be formed as slats. The main feature of this process is the transition from one unit of measurement (for rough-sawn stocks there is only volume) to two - pieces and volume (for each lamella there is fixed length, width and thickness). Accretion operations are recorded in the journal and entered into the system in the form of relevant document (Fig. 5).

| Act processing RSS (lengthening) |                                  |              |      |    |              |      |                                    |  |            |        |    |      |       |              |        |            |       |     |        |
|----------------------------------|----------------------------------|--------------|------|----|--------------|------|------------------------------------|--|------------|--------|----|------|-------|--------------|--------|------------|-------|-----|--------|
|                                  | RSS                              |              | Size |    | Wood species |      | Q-tx<br>#3                         |  | RSS length |        |    | Size |       | Wood species |        | Q-tx<br>#3 |       | Dif | % lost |
| Nr                               | Name                             | Series       | W    | T  |              |      |                                    |  | Name       | Series | L  | W    | T     |              |        |            |       |     |        |
| 1                                | RSS Oak 45x27                    | 91/06/17/003 | #5   | 27 | Oak          | 8,30 | RSS length Oak<br>91/06/17/003#P   |  |            | 1500   | 45 | 27   | Oak   | 197          | 6,2591 | 8,1429     | 26,18 |     |        |
| 2                                | RSS Beech 43x33                  | 91/06/17/005 | #3   | 23 | Beech        | 8,36 | RSS length Beech<br>91/06/17/005#P |  |            | 870    | 43 | 23   | Beech | 98           | 6,0517 | 8,0093     | 15,00 |     |        |
| 3                                | RSS Birch 47x33                  | 91/06/17/003 | #7   | 23 | Beech        | 8,45 | RSS length Beech<br>91/06/17/003#P |  |            | 870    | 47 | 23   | Beech | 452          | 6,249  | 6,0251     | 0,08  |     |        |
| 4                                | RSS Beech 50x24                  | 91/06/17/004 | #0   | 24 | Beech        | 8,20 | RSS length Beech<br>91/06/17/004#P |  |            | 1420   | 50 | 24   | Beech | 95           | 6,3619 | 8,0381     | 15,05 |     |        |
| 5                                | RSS Beech 45x31                  | 91/06/17/008 | #5   | 31 | Beech        | 8,39 | RSS length Beech<br>91/06/17/008#P |  |            | 850    | 45 | 31   | Beech | 271          | 6,3214 | 8,0636     | 17,59 |     |        |
| 6                                | Act processing RSS (lengthening) |              |      |    |              |      |                                    |  |            |        |    |      |       |              |        |            |       |     |        |

Fig. 5. Accretion accounting

To eliminate irregularities and defects which were formed during the seam, especially in the field of bonding, each adherent lamella is calibrated on the four-sided machines. Action of the machine is recorded in the journal and entered into the system (Fig. 6).

| 02. Act processing 4-side - A2 |                  |             |           |            |      |             |      |           |                  |             |      |        |    |      |      |        |        |      |
|--------------------------------|------------------|-------------|-----------|------------|------|-------------|------|-----------|------------------|-------------|------|--------|----|------|------|--------|--------|------|
|                                |                  | File        |           | Edit       |      | View        |      | Print     |                  | Insert      |      | Help   |    |      |      |        |        |      |
|                                |                  | Leave       |           | Print      |      | Close       |      |           |                  |             |      |        |    |      |      |        |        |      |
| 02. Act processing 4-side      |                  | date        |           | 01/31/2017 |      |             |      |           |                  |             |      |        |    |      |      |        |        |      |
|                                |                  |             |           |            |      |             |      |           |                  |             |      |        |    |      |      |        |        |      |
| No                             | HS-5 length      | Size        | Wood spec | Q'ty       | Q'ty | HS-5 length | Size | Wood spec | Q'ty             | Q'ty        | Dif  | % lost |    |      |      |        |        |      |
|                                | Name             | Series      | L         | W          | T    |             | Name | Series    | L                | W           | T    |        |    |      |      |        |        |      |
| 1                              | RSS length Beach | 03/12/17804 | 1400      | 35         | 24   | Beed        | 677  | 03/04/    | RSS length Beach | 03/12/17804 | 1400 | 30     | 23 | Beed | 677  | 0.0008 | 1.1063 | 1785 |
| 2                              | RSS length Beach | 03/12/17804 | 1400      | 35         | 24   | Beed        | 524  | 04/02/    | RSS length Beach | 03/12/17804 | 1400 | 30     | 23 | Beed | 524  | 0.3795 | 0.0002 | 1734 |
| 3                              | RSS length Beach | 03/12/17804 | 1370      | 47         | 25   | Beed        | 24   | 03/04/    | RSS length Beach | 03/12/17804 | 1370 | 43     | 19 | Beed | 24   | 0.0042 | 0.0222 | 3181 |
| 4                              | RSS length Beach | 03/12/17804 | 1400      | 35         | 22   | Beed        | 236  | 03/07/    | RSS length Beach | 03/12/17804 | 1400 | 37     | 19 | Beed | 236  | 0.1664 | 0.0012 | 2644 |
| 5                              | RSS length Beach | 03/12/17804 | 1370      | 39         | 23   | Beed        | 407D | 03/17/    | RSS length Beach | 03/12/17804 | 1370 | 37     | 19 | Beed | 407D | 0.0008 | 0.0002 | 2386 |
| 6                              | RSS length Beach | 03/12/17804 | 1370      | 32         | 22   | Beed        | 64   | 03/04/    | RSS length Beach | 03/12/17804 | 1370 | 27     | 19 | Beed | 64   | 0.0052 | 0.0219 | 2634 |
| 7                              | RSS length Beach | 03/12/17804 | 1370      | 32         | 23   | Beed        | 125  | 03/05/    | RSS length Beach | 03/12/17804 | 1370 | 27     | 19 | Beed | 125  | 0.0001 | 0.0049 | 1742 |
| 8                              | RSS length Beach | 03/12/17804 | 1300      | 35         | 21   | Beed        | 309  | 03/04/    | RSS length Beach | 03/12/17804 | 1300 | 30     | 23 | Beed | 309  | 0.2817 | 0.0013 | 1742 |

Fig. 6. Accounting of four-sided machine actions

This document allows you to bring data about size and number of lamellas before and after machine processing. The account of these units is very important, since the passage of the described operations occurs most technological material loss after roughing. Both documents contain field "difference", which has information about the material loss in  $m^3$  at each position. This specification allows providing clear operational supervision of the units, identifying and eliminating variations in the equipment or workers. Incorrect settings of machines or failures of the results occur significant excess of process loss that causes the increase of the finished product cost.

The norms of technological losses of raw materials are established for the data processing sections. The norm for accretion line is up to 10% by volume of rough-sawn stocks in overhauls and for four-sided machine - up to 20%. Monitoring of compliance with these rules can be carried out either immediately, right when entering data of nodes to systems, or aggregated over time using appropriate information reporting system (Fig. 7).

| Порядок | Дата | ИД документа | ЧМВ   | Объем ЧМВ     | Ламель | К-тв ламелей                     | Объем ламелей | Объем  | Чет. втрат |
|---------|------|--------------|-------|---------------|--------|----------------------------------|---------------|--------|------------|
| 2       | бум  | 01.06.2017   | 11006 | ЧМВ бум 47x25 | 64,28  |                                  |               | 58,46  | 9,13%      |
| 3       | бум  | 01.06.2017   | 11006 | ЧМВ бум 47x25 | 0,54   | Ламель зроцника бух 20/70x47x25  | 189           | 0,4596 | 0,0804     |
| 4       | бум  | 01.06.2017   | 11006 | ЧМВ бум 47x25 | 0,97   | Ламель зроцника бух 20/70x47x25  | 337           | 0,8196 | 0,1504     |
| 5       | бум  | 01.06.2017   | 11006 | ЧМВ бум 47x25 | 0,37   | Ламель зроцника бух 20/70x47x25  | 103           | 0,2591 | 0,0469     |
| 6       | бум  | 01.06.2017   | 11006 | ЧМВ бум 47x25 | 0,22   | Ламель зроцника бух 20/70x47x25  | 50            | 0,1251 | 0,0231     |
| 7       | бум  | 02.06.2017   | 11029 | ЧМВ бум 50x27 | 1,00   | Ламель зроцника бух 210/50x50x27 | 294           | 0,8135 | 0,1663     |
| 8       | бум  | 02.06.2017   | 11029 | ЧМВ бум 47x26 | 0,38   | Ламель зроцника бух 150/47x26    | 176           | 0,3236 | 0,0574     |
| 9       | бум  | 02.06.2017   | 11029 | ЧМВ бум 50x27 | 0,77   | Ламель зроцника бух 150/50x50x27 | 344           | 0,6966 | 0,1274     |
| 10      | бум  | 02.06.2017   | 11029 | ЧМВ бум 50x27 | 0,28   | Ламель зроцника бух 150/50x50x27 | 99            | 0,2539 | 0,0261     |
| 11      | бум  | 02.06.2017   | 11029 | ЧМВ бум 48x22 | 0,29   | Ламель зроцника бух 137/48x48x22 | 199           | 0,2749 | 0,0151     |
| 12      | бум  | 03.06.2017   | 11030 | ЧМВ бум 47x26 | 0,15   | Ламель зроцника бух 150/47x26    | 65            | 0,1191 | 0,0309     |
| 13      | бум  | 03.06.2017   | 11030 | ЧМВ бум 48x22 | 0,02   | Ламель зроцника бух 137/48x48x22 | 40            | 0,0518 | -0,0318    |
| 14      | бум  | 04.06.2017   | 11063 | ЧМВ бум 40x30 | 0,12   | Ламель зроцника бух 137/40x30x20 | 112           | 0,1228 | -0,0028    |
| 15      | бум  | 07.06.2017   | 11125 | ЧМВ бум 50x25 | 0,65   | Ламель зроцника бух 20/70x50x25  | 175           | 0,4709 | 0,1704     |
| 16      | бум  | 07.06.2017   | 11125 | ЧМВ бум 47x24 | 0,37   | Ламель зроцника бух 20/70x50x25  | 117           | 0,3037 | 0,0687     |
| 17      | бум  | 07.06.2017   | 11125 | ЧМВ бум 47x24 | 0,29   | Ламель зроцника бух 20/70x47x25  | 94            | 0,2296 | 0,0614     |

Fig. 7. Analysis of technological losses in lamella accretion

The system allows you to export reports to MS Excel for more detailed analysis and graphical display of results. Consolidated total analysis provides a more objective result, since the features of the process cause some uncertainty analysis for each position.

Fig. 8 Analysis of technology losses in four-sided machine

A similar analysis can be performed on the action of four-sided machine (Fig. 8).  
Derived final assemblies have fair size and can be used for manufacturing furniture panels.

Manufacturing furniture panels is a starting point of production of the finished product. The rough piece must pass a certain number of manufacturing operations to get finish product.

#### IV. CONCLUSION

Companies of different profiles, regardless of the size of the business, switch to automated computer systems that help to conduct business, monitor and plan it. Efficiency from this noticeably wins, especially in terms of saving time and money. The basis of automation in commercial organizations is ERP systems.

ERP systems are implemented with the purpose of uniting all the sectors of the company and all necessary functions in one management system which is to meet the requirements of all these sectors. The development of such a system is not an easy task.

Implementing of ERP systems by Ukrainian enterprises has special value as the tasks of the economic growth, increasing of production quality and competitiveness, becoming a member of the world market are most topical. But there are significant disadvantages of ERP implementation by Ukrainian SMEs including its high cost that is a special issue in crisis period.

We have described automation of manufacturing enterprises as it is a complex multistage process and the management information system of suchlike enterprise should include a number of modules majoring in certain areas.

We have provided alternative approach of above described automation the implementation of so-called scalable information systems-designers. Good examples of such systems are Accent 2 and 7.4. These systems are developed by Ukrainian IT-companies and enable complex automation of enterprises of any kind of activity. A typical set of functional requirements provides management and financial accounting, but open source framework allows coding modules for complex automation. Implementation of A2 and Accent systems is achievable in stages, progressively covering all areas of a company. In addition, implementation of such systems at the enterprise of wood processing industry may decrease its pollution and have good influence on ecology in the region.

#### REFERENCES

- [1] T. Kalling, "ERP systems and the strategic management processes that lead to competitive advantage", *Information Resources Management Journal*, Vol. 16, Iss. 4, pp. 46-67, 2003.
- [2] P. Azevedo, M. Romaro, E. Rebelo, "Advantages, Limitations and Solutions in the Use of ERP Systems (Enterprise Resource Planning) – A Case Study in the Hospitality Industry", *Procedia Technology*, Vol. 5, pp. 264-272, 2012.
- [3] L. Pang, Manager's Guide to Enterprise Resource Planning (ERP) Systems, *Information Systems Control Journal*, Vol. 4, pp. 47-52, 2001.
- [4] T. Davenport, "Mission Critical: Realizing the Promise of Enterprise Systems". Boston, Massachusetts: *Harvard Business School Press*, 2000.
- [5] S. Alshawi, S. Themistocleous, M. & R. Almadani, "Integrating diverse ERP Systems: a case study", *The Journal of Enterprise Information Management*, Vol. 17 (6), pp. 454-462, 2004,
- [6] M. Themistocleous, Z. Irani, R. O'Keefe, & R. Paul. ERP Problems and Application Integration Issues, in *proceedings of the 34th Hawaii International Conference on System Sciences: An Empirical Survey (HICSS-34)*, Vol. 9, pp. 1-10, 2001.
- [7] C. Soh, S. Kien, & J. Tay-Yap, "Cultural Fits and Misfits: Is ERP a Universal solution?", *Communications of ACM*, Vol. 43 (4), pp. 47-51, 2000.
- [8] M. Sumner. "Critical Success Factors in Enterprise Wide Information Management Systems Projects" in *proceedings of the 1999 ACM SIGCPR Conference on Computer Personnel Research*, pp. 297-303, 1999.
- [9] K. Murphy, & S. Simon. Intangible benefits valuation in ERP projects", *Information Systems Journal*, Vol. 12, 4, pp. 301-320, 2002.
- [10] Live Business, TOP 10 ERP solutions for Ukraine, 2017 [Online]. Available: <http://www.livebusiness.com.ua/tools/erp/> (in Russian)
- [11] Vasyl Grygoriv, Andrii Verstiak, Mariia Grygoriv, Svyatoslav Ishchenko, "Special aspects of Enterprise Resource Planning solutions in Ukraine", in *Proceedings 4th International Conference on Application and Communication xTechnology and Statistics in Economy and Education (ICAICTSEE – 2014)*, Sofia, Bulgaria, October 24-25, 2014, pp. 380-386.
- [12] Panorama Consulting Clash of the Titans 2017, 2017 [Online]. Available: <http://www.naveworld.com.sg/panorma-consulting-ranked-microsoft-dynamics-number-2-spot-top-erp-market-share-ranking>
- [13] Top five takeaways from the 2017 Clash of the Titans Report, 2017, [Online]. Available: <https://www.panorama-consulting.com/top-five-takeaways-2017-clash-titans-report>
- [14] IDC Homepage, [Online]. Available: <https://www.idc.com>, last accessed 2018/01/03
- [15] Ukraine imposes sanctions against 1C, Parus, ABBYY, Softline Group, Galatiki Center and other Russian IT companies, 2017 [Online]. Available: <http://en.interfax.com.ua/news/economic/421861.html>
- [16] Some dealers of 1C software in Ukraine placed on sanction list - SBU, 2017 [Online]. Available: <http://en.interfax.com.ua/news/general/418609.html>
- [17] Accent mainpage, 2017 [Online]. Available: <http://www.accent.ua/> (in Russian)
- [18] Accent as 1C alternative, 2017 [Online]. Available: <http://www.accent.ua/index.php/pages/ob-aktsente/alternativa-1s> (in Russian)

# Integration of Information Systems in Mergers and Acquisition of companies

Zora Říhová

Department of Applied Informatics, University of South Bohemia, CZECH REPUBLIC, Ceske Budejovice, email: zrihova@prf.jcu.cz

**Abstract:** The paper focuses on the issues of Mergers and Acquisitions (M&A) in case of integration of information systems (IS). Surveys show that in the area of IS it is extremely difficult to achieve success and it is important to do a very good IS analysis - requirements for the target functionality of the IS.

**Keywords:** Mergers and Acquisitions (M&A), information system, integration, analysis of IS, migration of data

## I. INTRODUCTION

Motto: "The Steering Committee on the issue of mergers usually spends more time choosing a new logo than dealing with the problems of IT and IS." D.Brown [3].

This paper is intended for the section Information and analytical support of economic activities. It deals with a very common issue, which in concrete terms is very difficult and not always successful.

One of the major economic activities that affects many companies throughout their life is the merger and acquisition. This is a one-time activity but it has a major impact on the companies. Therefore, a deep analysis and setting of clear objectives to be achieved by this process is a very essential prerequisite. This paper focuses on one part of this process which is the way of integrating information systems .

The integration involves several different ways of IS integration, based on personal project experience from the years 2002-2015[7].

In practice, we distinguish these basic types of M&A [4]:

*Vertical merger* is a combination of businesses at different levels of the supply chain. This is a connection to a vendor (extension to sources), or to customers (spreading to the customer). For example, the glass manufacturer will contact the installer.

*Horizontal merger* is a combination of two or more companies operating in the same industry and the market at the same level of the supply chain. The merger of two competing businesses, where, for example, a smaller company merges with a larger one. An example could be the merger of two district bakeries.

*Congeneric merger* means a merger of companies from related business unions that do not compete with each other but complement each other. A classic example is the connection of a computer-centric company and a software developer.

*Conglomerate merger* is the type of merger involving companies from unrelated business sectors. For example, making toys with a construction company.

In Table 1 is the overview of M&A types and their impact on IS by strategy types.

TABLE 1. REQUIREMENTS FOR IS INTEGRATION BY STRATEGY TYPES [8]

| M&A Strategy   | Model           | Requirements for IS integration   |
|----------------|-----------------|---|
| Vertical       | Unification     | Real-time exchange and sharing of data, highly integrated business processes                                  |
|                | Coordination    | Exchange and data sharing,  |
| Horizontal     | Unification     | Fully integrated company, data exchange and sharing   |
|                | Coordination    | Exchange and sharing of customer information, centralized management, autonomous operation of business units  |
| Congeneric     | Replication     | Centralized and standardized business processes, aggregation of information                                   |
|                | Coordination    | Aggregation of information, sharing of customer / market information, autonomous operations of business units |
| Conglomerative | Diversification | Independent transactions, Shared Infrastructure/Service, Autonomous Business Unit Operations                  |
|                | Diversification | Autonomous operations of individual business units, independent transactions                                  |

A somewhat related issue is the efficiency of of integration of information systems which is, however, not dealt with in this paper. Lipaj and Davidaciene[9] discuss how integration of information systems can increase the economic indicators of businesses, together with work productivity and efficiency. The paper by Yuldasheva & Doronina [10] is devoted to the task of developing efficient means to analyze functioning of an

information system of the cyclic type based on determining the integral performance criterion. The criterion is developed based on the analysis of all the processes occurring in the information system (IS) of a university. Also they consider certain qualitative and quantitative parameters affecting the integral index of efficiency.

## II. GOALS OF INTEGRATION

The reason for mergers and acquisitions is to achieve financial success and power, savings, higher shareholder value: so it is mainly about getting a positive synergistic effect. From the literature [1] it is known that the issue of ICT is not often seen as crucial in the course of trade talks. On the basis of a survey which was conducted in 80 companies in cooperation with 20 CEOs, 20 CFOs, 20 IT and HR Directors, it was found that Information technology occupies the largest share of time and this is the reason to address this issue.

The goals of the integration of IS in the process of M&A is [7]:

- to support the achievement of the objectives set in combining companies
- to standardize processes
- to standardize the methodology for the calculation of basic indicators
- to consolidate information technology without disruption of companies
- to prepare and implement synergies into the cost of IT
- to create conditions for the effective use of IS.

Therefore, it is necessary to analyze the situation of IS.

The above described points are given as an example which factors are essential to be taken into account in the course of an integration process.

## III. APPROACHES TO INTEGRATION

The decision about the final IS has to be taken by the management. Prior to this decision it is necessary to conduct a feasibility study.

During the feasibility study the following tasks have to be performed and documented:

- Detailed analysis of the existing functionalities and processes
- Gap analysis
- Gap classification (relevance)

- Alignment and documentation of the functional specifications
- Performance criteria and requirements
- Proposals of the target IT architecture and processes
- Cost estimation including test efforts
- Implementation and testing plan with special regard of the 'smooth' outphasing of the current solution (e.g.analyze if Big Bang is possible)
- Risk analysis and mitigation actions

Based on the outcome of the feasibility study the IT staff makes proposals and recommendations for implementation possibilities: [7]:

- a) unification - highly integrated business processes
- b) coordination – exchange and sharing of information
- c) replication – aggregation of information on a higher level
- d) diversification – only a few shared data (except financial information).

For the merger of Information Systems following different solutions can be recognized:

- a) leaving existing information systems and creating a basic/critical connection of some selected functions or system specific *tady něco chybí*
- b) moving the main functionality in a common system
- c) migration of one used information system to the other existing information system
- d) migration of all companies in the new joining.

Selection of variants in a) through e) depends on the selected IS investments, but primarily on the relationship to the production and the model of M&A.

Integration always requires financial, human, material and time costs. It is important that consideration should be given to the necessary adjustments.

M&A in the field of IS is the most complex task for IT staff s is a alignment of strategy, gap analysis systems, data analysis, data migration, etc.

Factors (Fig.1) with relevance of 77-50 refer mainly to owners and company management, possibly to several experts. But relevance 48 (this is also a very high figure) is the integration of the IS as a technological activity, which can be attended by up to hundreds of people who analyze, elaborate, search for uniqueness and requirements. This requires not only the involvement of many people, but also time-consuming work.

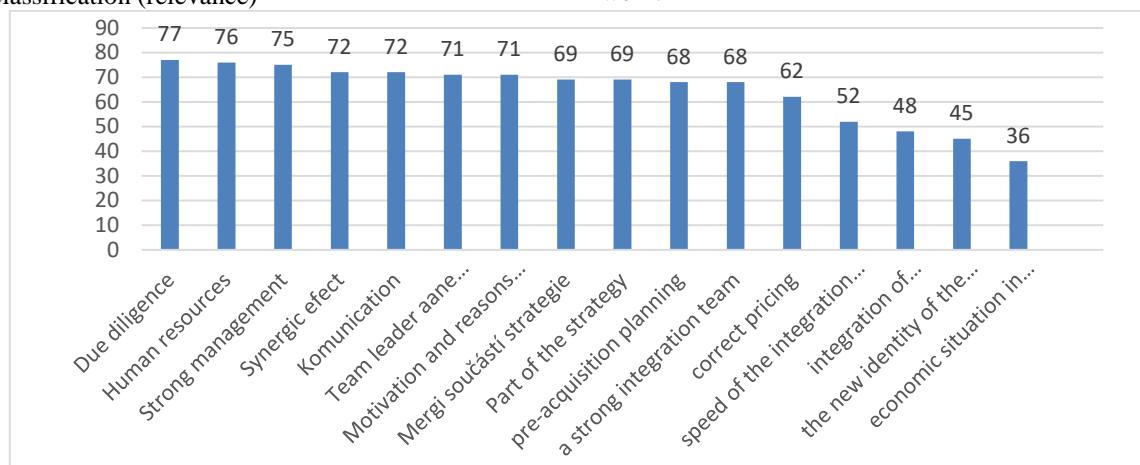


Fig. 1. The relevance of individual factors in relation to the success of M&A [63].

Experience shows that the steering committee often does not realize that this interferes all areas of the company, and thus it differs from integration activities in other area. The same is

true for setting the time for this integration. During the M&A all information systems have to work in parallel.

Similar results are from research Factors of Failure [2].

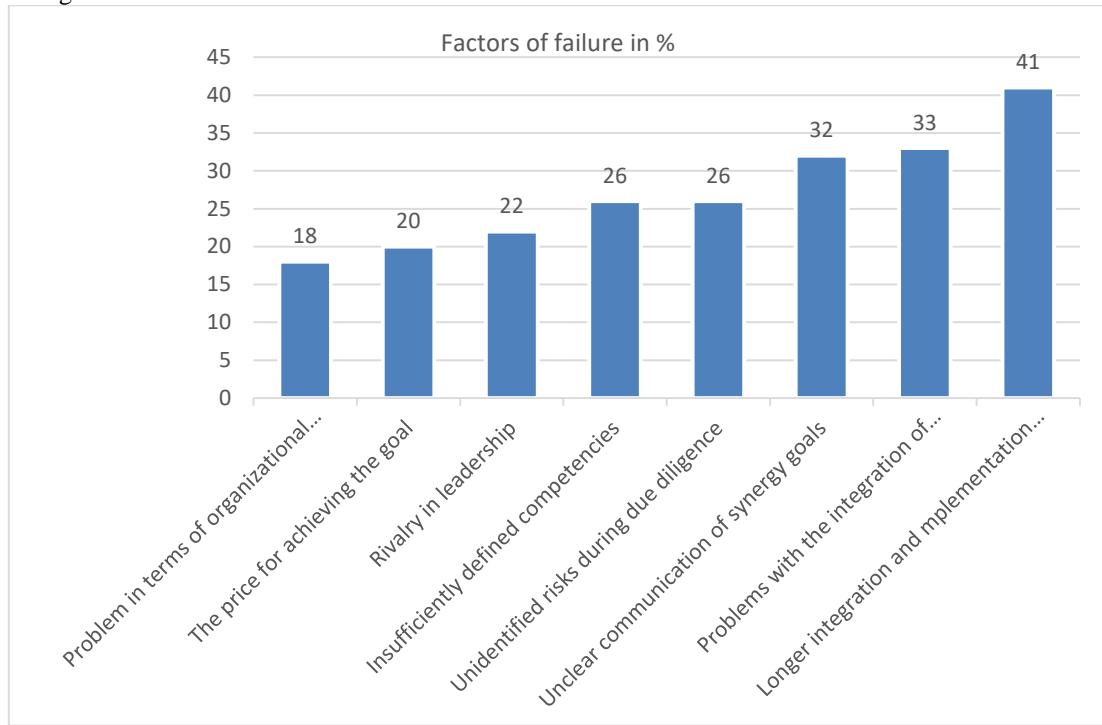


Fig.2. Factors of failure [2].

Problems with IS are very high 41% and it is necessary to solve it. The another text deals with the causes of IS issues.

#### IV. ROOT CAUSES OF PROBLEMS

From the praxis it is known that questions about IS during trade negotiations in M&A are rarely seen as crucial.

Failure to achieve the expected results is 70% the result of problems in the field of ICT, as shown in the study which was done by Hugh Craig-Halkett, who examined the CEO and IT due diligence in Great Britain and the USA [3].

One study done by Accenture [1] showed that more than 75% of executives are unaware of the critical role of IT in M&A.

The study [6] showed that more than 75% from 125 companies in the survey had problems in the integration of ICT, particularly the incompatibility of information systems.

These problems were in unconnected systems and processes, data redundancy, incompatibility, slow integration and thus lost business opportunities, which are often discovered only after signing the M & A contract.

The root causes of problems in M&A process are:

- Insufficient gap analysis** before signing the contract.  
- setting of clear objectives, finding specification and uniqueness in all IS and requirements to the target IS functions.
- Analysis of the Integration Process** - this is a unique / individual process for which there is no methodology available and, therefore, has to be preceded by a thorough analysis of the intent and the procedure. It turns out that corporate culture plays a significant role in integration, but unfortunately IT staff is not sufficiently analyzed and perceived.

c) **Migration and consolidation of data.** This area is often underestimated and solved often after signing the M&A contract. It should be clear whether data for continuous production (joint planning and production, bank customer service, etc.) is being processed and consolidated or information is being processed (data is being evaluated), and in this case a more loose integration schedule, different data structure and different processing methods, inaccessible documentation not only before the integration but also the course of integration, which can not be of a technical nature. It is necessary to define data and their models, link to processes and individual transformation points in the enterprise.

d) **Non-compliance framework to comply with the data governance.** Compliance can help integrate workflows and accelerate the achievement of integration goals (accessibility, manipulation, accuracy and completeness, consistency, controllability and security). This area is fully responsible and responsible for ICT staff and quality project management of the M&A.

e) **Project management of the integration process.** Mastering the project is not a matter of course and has to start with the correctly selected team including determining responsibility for this integration. Then, of course, it is necessary to set and follow the project procedures (carefully elaborate a project management plan, project goals, determine risks, think real timetable, etc.). An important role, often decisive, is the personality of the project manager. It also depends on the team building, cooperation, organizational and factual management.

f) **Underestimation of the time required for the integration of IS systems.** Schedules usually set times for the ideal state of progress and the course of M&A. It is not just

about the speed at which integration takes place depending on the need for data integration - and therefore in terms of meeting the planned ICT integration timetable, but also from the point of view of starting this integration (some M&A starts the integration process even after a few years after signing the contract). In this case, the solution is the decision of the steering committee, based on the recommendations of the ICT staff, when sufficient timetable clearance or appropriate capacity to implement the overall M&A.

Further successful management of the M&A process in area of IS is a complex process of tasks in which there is a potential risk of occurrence of a number of errors and problems. The other related problems can be classified into the following seven categories:

- Inappropriate planning
- Unrealistic expectation of the project
- Bad time distribution
- Inefficient use of resources
- Miscommunication
- The absence of sophisticated business tools
- Difficult access to quality information.

The occurrence of each of these problems can potentially also lead to a failure, even if the most effective opportunity for merger or acquisition is to be seen.

## V. CONCLUSION

Problems with IS and its integration together with the insufficient data analysis of associated companies is often the cause of failure in the resulting M&A.

This paper generalized the specific findings of the five mergers IS of large multinational companies.

IS is the nervous system of the company and its interconnection to another nervous system is a complex process.

So it is therefore important for success M&A of IS:

- Accurate and thorough definition of objectives and scope of the project
- Establishing sufficiently realistic plans in relation to resource capacity
- Defining basic competencies (roles and powers),
- Ensuring sufficient expertise in data migration and project management
- Sufficient awareness of the ongoing and ongoing project of all stakeholders
- Collect all available documentation describing source and target systems and data
- Sufficient understanding of functional and other requirements
- Reduction of complex dependencies on related projects
- Identification of potential problems and risks.

Therefore, an analysis of the IS, involving a large number of people (users, consultants, experts, etc.), and individual information is based on data and therefore data analysis and migration are important components of successful M&A

implementation. It is important for this to be known by the steering committee of the participating companies.

Integration of IT often comes up after the M&A contract is signed and then this area is resolved smoothly. It is in the pre-contractual process that possible future problems of integration of IT are usually not solved. This can lead to unforeseen additional costs and time slippages in the schedule.

If the assessment options compatibility of the IT were engaged before signing the contract, then a real plan for the integration along with an assessment of the risks and possibilities for risk prevention could be developed.

One of the most important motives for mergers and acquisitions is to create a more stable and efficient company with significant and competitive market activity. However, even after a successful merger, maintaining the competitiveness of a company can be a difficult matter. One important way is consistent attention to data and information systems.

## REFERENCES

- [1] ACCENTURE, „Keys to the kingdom: How an integrated IT capability can increase your odds of M&A success“,. Available from: [www.accenture.com](http://www.accenture.com).2002, 2002
- [2] AON HEWITT, 2011 “Culture integration in M&A”, *Consulting M&A Solutions: Survey Findings*. Available from: [www.aon.com](http://www.aon.com)
- [3] D. Brown, “Don’t overlook IT in the Merger,” *Computer Weekly*, 2001
- [4] E. Kislingerová, ” Manažerské finance.” *Praha C.H. Beck 2007. ISBN 978-80- 7179-903*
- [5] L. MILNE, “Success in mergers and acquisitions”, Milnerllp 2010. Available from: [www.milnerllp.com](http://www.milnerllp.com)
- [6] S. Konkolski, „Strukturalizace a analýza faktorů ovlivňujících efektivnost fúzí a akvizic“. *Hranice, 2011. Dizertační práce*.
- [7] Z.Řihová, “The Role of ICT in Mergers and Acquisitions”. In: *Liberecké informatické fórum*. Liberec, 08.11.2012 – 09.11.2012. Liberec : Technická univerzita v Liberci, 2012, pp. 99–105.
- [8] Z.Řihová, “Joining the company in terms of integration of information systems”, Prague, *Journal System Integration 4*, 2012
- [9] D. Lipaj, & Davidaviciene, V. “Influence of information systems on business performance“. *Mokslas: Lietuvos Ateitis, 5(1)*, 38. 2013.
- [10] T. Yuldasheva, & J. Doronina, “Analysis of the Integral Efficiency Indicator for Information Systems of the Cyclic Type Accounting for Weights“. *Journal of Data Analysis and Information Processing, 1(02)*, 9. 2013
- [11] I.Cubed - Critical Success Factors of CAD Data Migrations. Available from: <http://www.i-cubed.com/pdf/CAD-Migration-Success-Factors.pdf>, 2011

# IT Law: Implementation Problems and Development Prospects

# Cybercrimes, Cyber Law and Computer Programs for Security

Antonina Farion<sup>1</sup>, Valentyna Panasyuk<sup>2</sup>

1. Department of Economical Security and Financial Investigation, Ternopil National Economic University, UKRAINE, Ternopil, 46 A Mykulynetska str., email: [secretmail\\_antonina@ukr.net](mailto:secretmail_antonina@ukr.net)
2. Department of accounting in the industrial sphere, Ternopil National Economic University, UKRAINE, Ternopil, Peremohy Square 3, email: [Tina.panasjuk@gmail.com](mailto:Tina.panasjuk@gmail.com)

**Abstract:** In this document we describe the situation that was formed at the information market colligates with the increasing the level of cybercrimes. Law regulation of this sphere can't follow the development of information technology that exacerbates the problems of cybercrime. At the individuals' level cybercrime is associated with the using of pirated software: malicious people can access the user's personal date.

**Keywords:** information technology, cyberspace, intellectual property, cyber security, antivirus and protection.

## I. INTRODUCTION

Law and Information Technology are parallel objects and many scientists prove that they complement each other. A lot of lawyers complain that law is always running behind the process of developing information technology.

## II. THEORETICAL BASIS

R. M. Kamble underlines that information technology deals with information system, data storage, access, retrieval, analysis and intelligent decision making. Information technology refers to the creation, gathering, processing, storage, presentation and dissemination of information and also the processes and devices that enable all this to be done<sup>1</sup>. And computers become inalienable part of our life. Cybercrime is defined as crimes committed on the internet using the computer as either a tool or a targeted victim. Cybercrimes involve both the computer and the person behind it as victims; it just depends on which of the two is the main target<sup>2</sup>. So cyberspace spreads and become more dangerous because many people can be involved in it. Criminals roam freely in cyberspace than in other environment.

## III. PRACTICE

Cybercriminal activity is one of the biggest challenges that humanity will face in the next two decades<sup>3</sup>. It is predicted

that cybercrime will cost the world \$6 trillion annually by 2021. This increasing are based on hundreds of major media outlets, universities and colleges, senior government officials, associations, industry experts, the largest technology and cybersecurity companies, and cybercrime fighters globally (Fig.1).

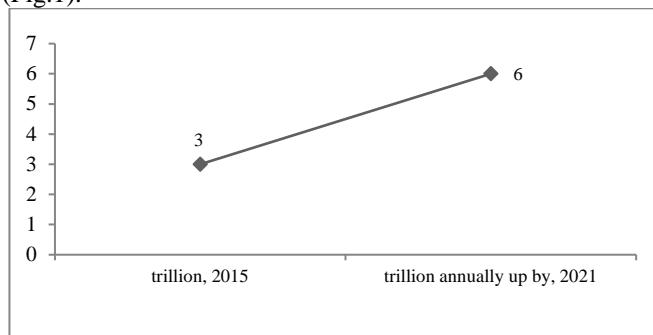


Fig. 1. Prediction for increasing of cybercrimes cost from 2015 annually up by 2021

It is direct connection of changing amount of internet users: 100000 in 1990 and 500 million people in 2013. These date rapidly changed (Fig. 2).

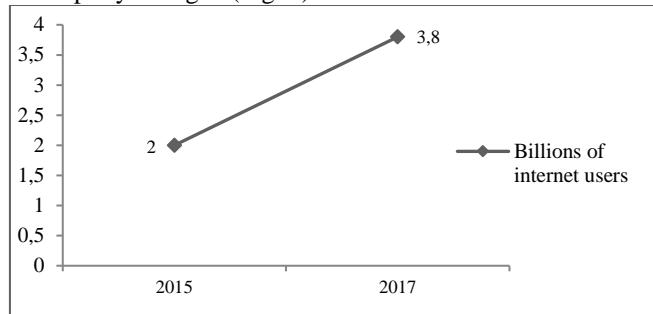


Fig. 2. Quantity of internet users changing (2015-2017)

Cybercrimes have unique structure that is connected with information technologies (Fig.3).

Many crimes that involve the use of cyber-technology are not genuine cybercrimes. Cyber-related crimes could be further divided into two sub-categories:

- cyber-exacerbated crimes;
- cyber-assisted crimes.

Crimes involving cybertechnology could be classified in one of three ways: cyber-specific crimes (genuine cybercrimes); cyber-exacerbated crimes; cyber-assisted crimes.

<sup>1</sup> R. M. Kamble. Cyber law and information technology. International Journal of Scientific & Engineering Research, Volume 4, Issue 5, May-2013

<sup>2</sup>Computer Crime Research Center. Cybercrime definition. Electronic access: <http://www.crime-research.org/articles/joseph06>

<sup>3</sup> Steve Morgan, Editor-in-Chief Cybersecurity Ventures. 2017 Cybercrime Report. Herjavec group. Electronic access: <https://cybersecurityventures.com/2015-wp/wp-content/uploads/2017/10/2017-Cybercrime-Report.pdf>

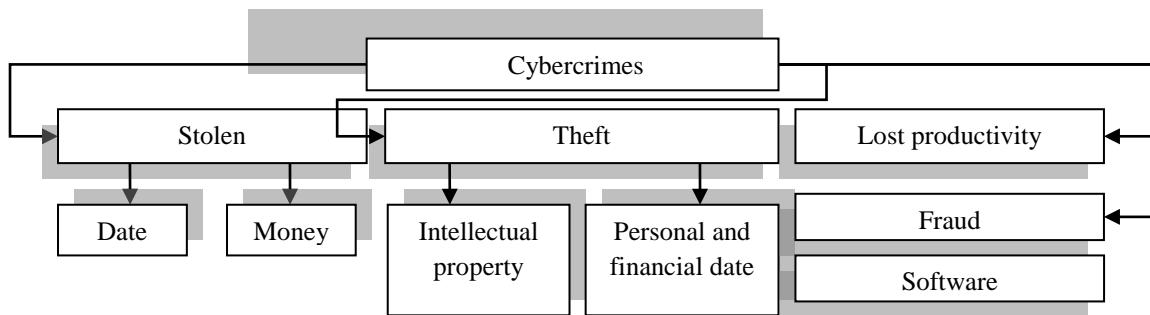


Fig. 3. Cybercrimes that connect with information technologies.

Like other kinds of crime, which historically grew in relation to population growth, cybercrimes grow in proportion to digital targets. And cybercrimes are more dangerous than the others because criminals can operate anonymously over the computer networks. The difference between crimes is the hackers steal intellectual property. Law that connects with cybercrimes must cover IT area (Fig. 4).

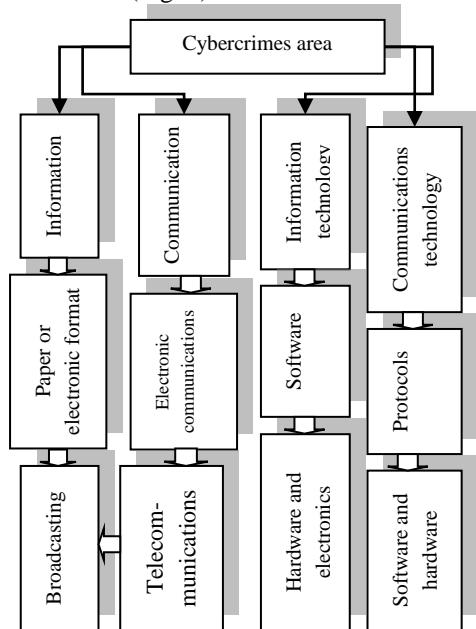


Fig. 4. Definition of cybercrimes area.

During last 20 years many security software were invited for electronic date protection because in the world's practice there is not the single law that can regulate all IT relations.

There is a field of law that comprises elements of various branches of the law<sup>4</sup> (Fig. 5).

But even these parts of law are not enough to control cyberspace. Cybercrimes develop more quickly than others crimes (Fig. 6) [1]-[6]. Many countries have very few laws addressing cybercrime.

- Love Bug Virus;
- VB script that spread via email and corrupted many different file types;
- FBI traced the virus to the Philippines.

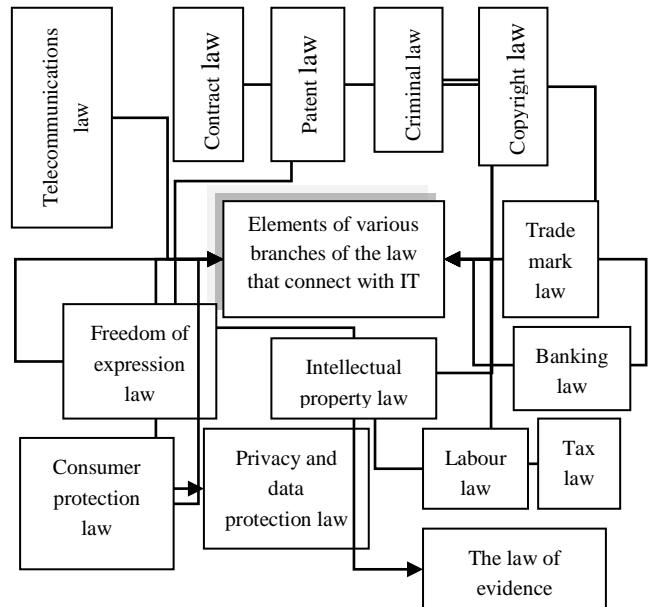


Fig. 5. Elements of various branches of the law that connect with IT for creation the unique law for protection internet users' property.

But can legislation stop cyber crime. Research shows that the costs of cyber crime for companies in financial services and utilities and energy have the highest annualized cost. The most expensive attacks are malicious insiders, denial of service and Web-based attacks [7]. In last 2017 year in the world the new kinds of cybercrimes appeared – machine learning accelerates social engineering attacks or cloud computing providers' infection. But the necessary sections in the law that provide security from cybercrimes are not adopted so quickly. So, cybersecurity is the main instrument in securing date from threats (Fig.7).

Many computer criminals have been company employees, who were formerly loyal and trustworthy and who did not necessarily possess great computer expertise. To prevent increasing in cybercrimes activity it is important to identify career criminals, including those involved in organized crime, who are now using cyberspace to conduct many of their criminal activities. Some cyber-related crimes can be carried out by professional's offenders and might be undetected because professional criminals do not typically make the same kinds of mistakes as hackers, who often tend to be amateurs.

<sup>4</sup> What is IT law, ICT law or Cyber law? Michalsons. Electronic access: <https://www.michalsons.com/blog/what-is-it-law-ict-law-or-cyber-law/286>

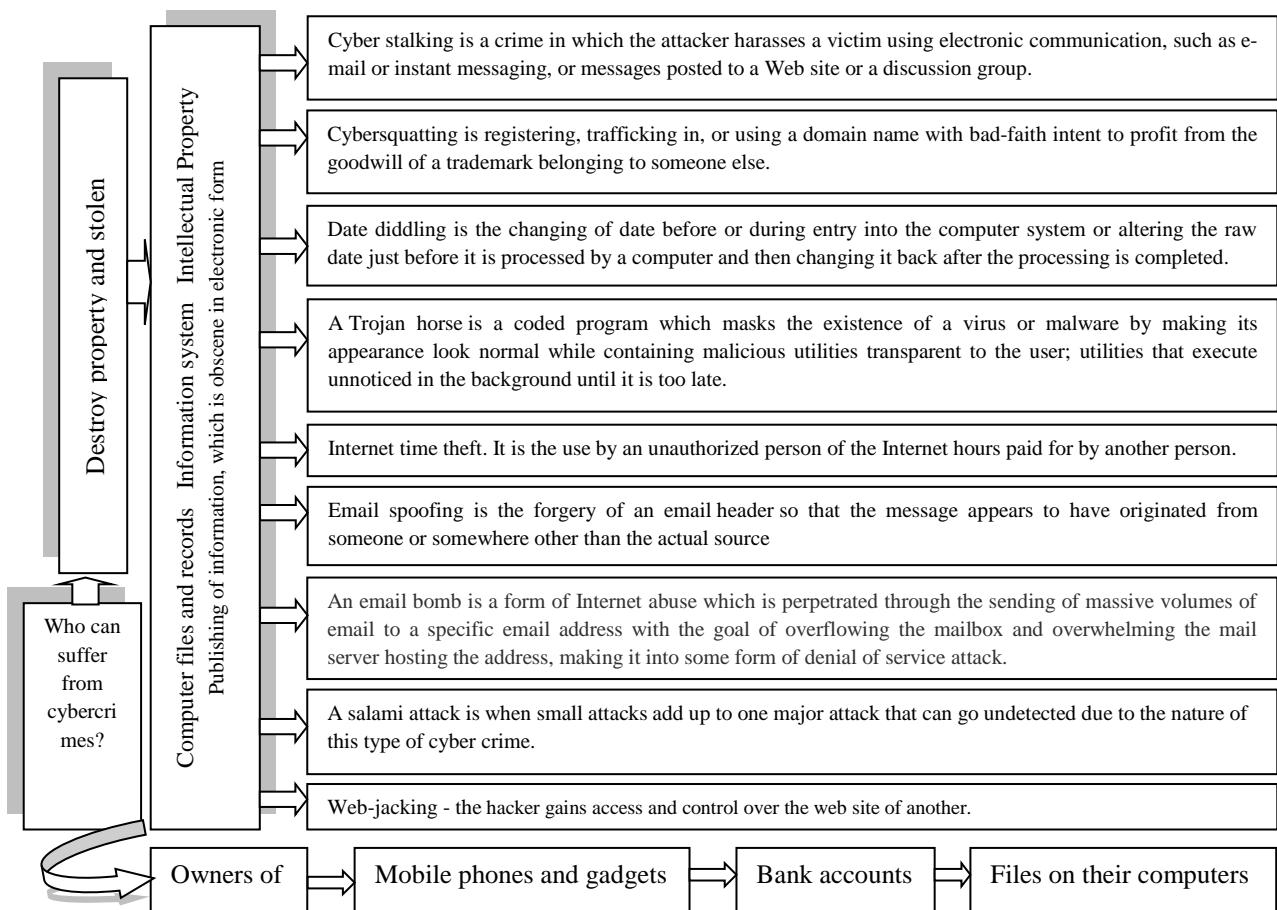


Fig. 6. Classification of Cyber Crimes and people who are affected by them.

Information age is so called because our life is codified by date: almost everything we do or buy, and everything we depend on, involves data and the technology that uses it. Cyber criminals are building so called “an army of things” that has the potential to impact the future of the digital economy [9]. Impact of a cyber attack could include substantial loss of revenue and margin, of valuable data, and of other company assets. Quantity of cybercriminals increases quickly around the world. Now cybercrimes are connected international serious organized crime groups, smaller-scale, domestic criminals and hacktivists.

Although the most serious threat comes, directly or indirectly, from international crime groups, the majority of cyber criminals have relatively low technical capability. Their attacks are increasingly enabled by the growing online criminal marketplace, which provides easy access to sophisticated and bespoke tools and expertise, allowing these less cyber criminals to exploit a wide range of vulnerabilities [10]. There is also situation when companies’ websites were subject to the criminal access of a customer records database, followed by a ransom demand asking for payment in exchange for the return of stolen data. The wearables are rapidly gaining popularity with smartwatches. Wearables are tracking all sorts of personal information including GPS location, blood pressure, heart rate, and anything else you feed them such as weight or diet. Such personally identifiable information could be used as a base to target you for spear-

phishing, or aid in identity theft. But the real opportunity is these devices linking to your smartphone, where phone numbers, more personally identifiable information, emails, web logins etc. could theoretically be compromised [10].

Cybercrime activity is spreading around the world. For decreasing the cybercrimes in Europe, Cooperation Group, the Commission, the European Union Agency for Network and Information Security should be established to support information security within the EU countries [12]. According to Directive (EU) 2016/1148 the certain sectors of the economy are already regulated or may be regulated in the future by sector-specific Union legal acts that include rules related to the security of network and information systems. Each Member State shall designate one or more national competent authorities on the security of network and information systems.

Member States shall ensure that digital service providers identify and take appropriate and proportionate technical and organizational measures to manage the risks posed to the security of network and information systems which they use in the context of offering services referred to in Annex III within the Union. Having regard to the state of the art, those measures shall ensure a level of security of network and information systems appropriate to the risk posed, and shall take into account the following elements: the security of systems and facilities, incident handling, business continuity management, monitoring, auditing and testing [12].

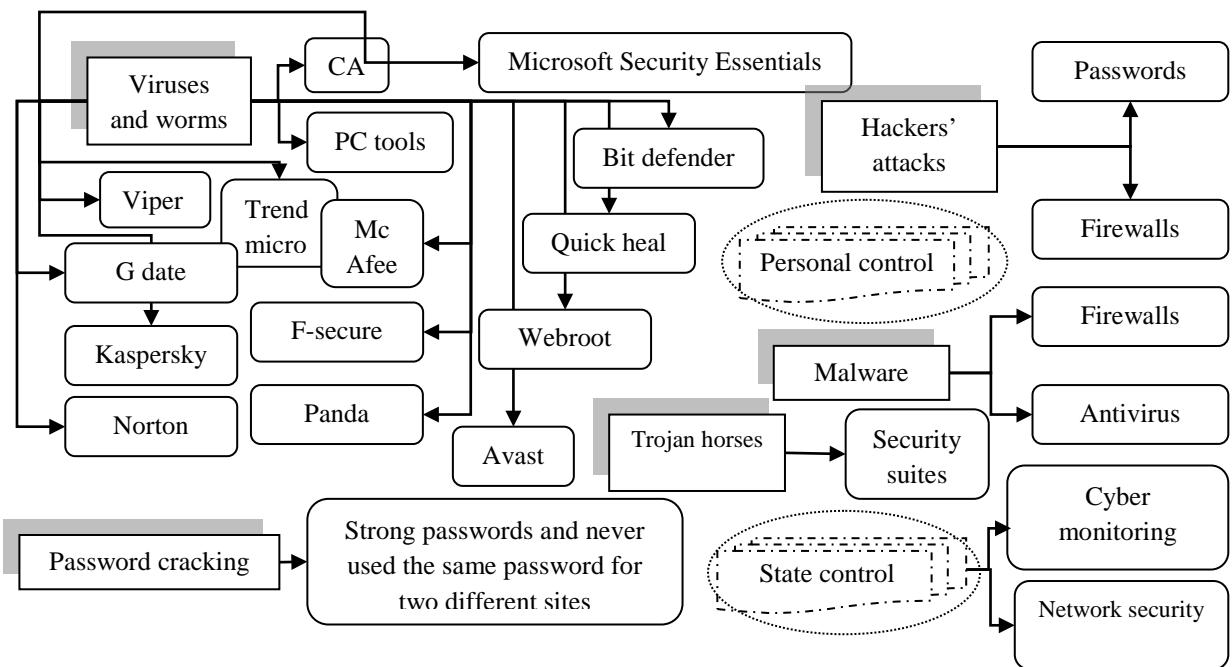


Fig. 7. Types of security in the network space.

The top industries at the greatest risk of cyber attack (Fig. 8) [11].

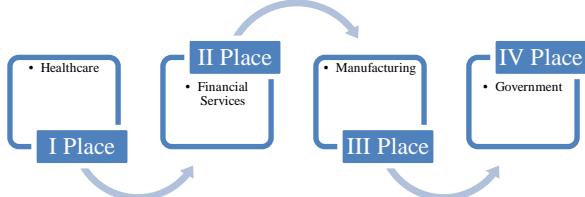


Fig. 8. The top 4 industries at the greatest risk of cyber attack.

#### IV. CONCLUSION

Cyber crimes are intrinsically challenging for business companies and governments. Security has to be developed quicker than types of cyber crimes because solutions that may have worked last year may not necessarily work this year or next.

#### REFERENCES

- [1] Cyber stalking. Available at: <http://searchsecurity.techtarget.com/definition/cyberstalking>
- [2] Cybersquatting. Available at: <http://searchmicroservices.techtarget.com/definition/cyber-squatting>
- [3] E-mail spoofing. Available at: <http://searchsecurity.techtarget.com/definition/email-spoofing>
- [4] What is a Trojan Horse Virus? - Definition, Examples & Removal Options. Available at: <https://study.com/academy/lesson/what-is-a-trojan-horse-virus-definition-examples-removal-options.html>
- [5] Aj. Maurya. What is a salami attack? Available at: <https://ajmaurya.wordpress.com/2014/03/27/what-is-a-salami-attack/>
- [6] Email Bomb. Electronic access: <https://www.techopedia.com/definition/1655/email-bomb>
- [7] Cost of cyber crime study. Insights on the security investments that make a difference. Independently conducted by Ponemon Institute LLC and jointly developed by Accenture. Available at: [https://www.accenture.com/t20170926T072837Z\\_w/us-en/acnmedia/PDF-61/Accenture-2017-CostCyberCrimeStudy.pdf](https://www.accenture.com/t20170926T072837Z_w/us-en/acnmedia/PDF-61/Accenture-2017-CostCyberCrimeStudy.pdf)
- [8] Cyber security. Available at: <https://www.slideshare.net/Siblu28/cyber-security-36922359>
- [9] Cyber criminals a growing threat to digital economy. Available at: <https://www.gtnews.com/2017/03/29/cyber-criminals-a-growing-threat-to-digital-economy/>
- [10] NCA Strategic Cyber Industry Group. Cyber Crime Assessment 2016. Need for a stronger law enforcement and business partnership to fight cyber crime. Available at: <http://www.nationalcrimeagency.gov.uk/publications/709-cyber-crime-assessment-2016/file>
- [11] 5 industries that top the hit list of cyber criminals in 2017. Available at: <http://www.infoguardsecurity.com/5-industries-top-hit-list-cyber-criminals-2017/>
- [12] EUR-lex. Directive (EU) 2016/1148 of the European Parliament and of the Council. Available at: <https://eur-lex.europa.eu/legal-content/EN/TXT/?toc=OJ%3AL%3A2016%3A194%3ATOCT&uri=uriserv%3AOJ.L.2016.194.01.0001.01.ENG>

# Incitement to Suicide with Social Networks and the Internet: Problems of Criminal Liability in Ukraine

Oksana Yaremko<sup>1</sup>, Serhiy Banakh<sup>2</sup>

1. Department of International Law, International Relations and Diplomacies, Ternopil National Economic University, UKRAINE, Ternopil, Mikulinetskaya str., 46-A, email: yaremko\_oksana@ukr.net

2. Faculty of Law, Ternopil National Economic University, UKRAINE, Ternopil, Mikulinetskaya str., 46-A, email: s.v.banakh@gmail.com

**Abstract:** It has been substantiated legal exposure of children on the internet due to the cases of incitement to suicide. It has been analyzed the problems of legislation and law enforcement practice of criminal responsibility for incitement to suicide with the use of internet. It has been proposed the ways of their improvement.

**Keywords:** internet, social networks, incitement to suicide, crime.

## I. INTRODUCTION

With the development of technical progress and the evolution of mankind, new, previously unknown crimes and ways to commit them appear. So, with the advent and development of the Internet there were new ways of committing not only theft or fraud, but also suicide. At the same time, a new way of incitement to suicide appeared in the world, along with an increase in the number of regular users of the Internet.

## II. THE PROBLEM OF SUICIDE IN UKRAINIAN SOCIETY

Ukraine is among the top ten countries with the highest suicide rate. The ratio of suicides and incitements to suicide is about one to twenty. At the same time, in Ukraine, from 2010 to 2015, only four sentences were imposed for incitement to suicide, that is, less than one sentence per year [1]. Normally, the law enforcement agencies stopped proceedings in suicide cases without finding traces of violence on the body of the dead. In practice, in the case of a person's death, the version of suicide is not properly investigated or ignored in general. Moreover, under current legal realities, Ukrainian legislation and practice of law enforcement turned out to be helpless ("unarmed") before a new method of suicide – incitement to suicide using social networks and the Internet.

Due to the high latency of suicide attempts and the absence of norms in the Criminal Code of Ukraine (hereinafter – the CC of Ukraine) until February 8, 2018 [2], which would give the possibility of prosecution for incitement to suicide through social networks and the Internet, to assess the actual situation and condition the problem was rather difficult before. For example, only through a private investigation of a relative of one deceased girl (who was subjected to suicide using the Internet) and through a journalistic investigation conducted at the request of the aforementioned relative, nine people were incited to suicide. The materials of this private investigation were transferred to law enforcement agencies. As a result, a criminal case was opened in Zaporizhzhya in September 2015 under Article 120 of the Criminal Code of

Ukraine in relation to Yevgen Krivosheyev. He was suspected of incitement to suicide through the Internet of nine young women, four of whom lived in the territory of the Russian Federation, and five women were the residents of Ukraine. Unfortunately, only one of the victims of Yevhen Krivosheyev, who lives on the territory of Ukraine managed to save [3].

Under this criminal proceedings, the investigators defined IP-address of the person who using the woman's name in the social network "VKontakte" looked for the future victims and drove them to suicide. The apartment of the suspect, Yevgen Krivosheyev, was searched and seized a computer from which the person suspected of nine victims incitement to suicide came to the Internet. According to the rescued victim of the crime, the suspect called himself a female name, presented himself as a nurse, gave advice and recommendations on how to commit suicide, helped to choose the place for hanging tools and its implementation, exhorted to do it together. The correspondence of the suspect with other victims (which, unfortunately, could not be saved) confirmed his methods of driving to suicide. It should be noted that after the death of the victims, the suspect wrote rather rude things to their relatives in social networks [4].

However, it was not possible to bring Yevgen Krivosheyev to criminal liability. Indeed, at the time of committing these crimes, i.e. incitement to suicide nine persons, there was no rule that would provide criminal penalties for this crime in the Criminal Code of Ukraine.

## III. LEGAL EXPOSURE OF CHILDREN ON THE INTERNET AND SOCIAL NETWORKS FROM THE IMPACT INCITEMENT TO SUICIDE

The lack of power of Ukrainian legislation and law enforcement officers to withstand the tendencies of growing crime on the Internet and the emergence of a new method of suicide is also evidenced by the terrible consequences that our society has suffered from the spread of social networks, the so-called "death groups".

So, in recent years, the number of suicides among children has increased in Ukraine. For example, in 2015 half a hundred children left their lives with the suicide. In this case, the boys among them were twice as many as girls. Commissioner of the President of Ukraine on the rights of the child Mykola Kuleba relates this to the development of "death groups" in the most popular social networks VKontakte, Facebook, Twitter and Instagram [5].

It is worth noting that at the end of December 2017, the Investigation Department of the National Police of Ukraine reported on the discovery of almost 1,000 "death groups",

600 of which had already been blocked, and the prevention of 10 suicide attempts. As a whole, there were almost 35 thousand signatories from Ukraine in these public associations. According to the opinion of the head of the National Police of Ukraine Serhiy Knyazev: "Children are most vulnerable to the negative influence of social networks especially of 13-17 years old who have problems in communicating with others." [6].

Meanwhile, in 2015 the mass media were widely speaking about the first series of childhood suicides, which were initiated by "death groups" from social networks where the victims were mostly children. Mass media actively talked about popularization among children of terrible "games", participation in which should cost them (children) their life. One of the deadly games is the Blue Whale. Symbols of this game are whales, which are chosen by the administrators of the "group of death" non-accidentally, because the whales are one of the few species of mammals, which can stop their lives by spontaneously throwing on the shore. Participants of the game distributed tasks (for example, cut their hands with a blade, taking all of this process on the video), where the last task was a suicide, shooting on the camera [7]. According to the evidences of the cyberpolicemen, the content of the shootings taken during the game on the camera of the facts of driving the children to suicide is very big money in the closed online forums.

Meanwhile, the Blue Whale is not the only danger of children and adolescents in social networks and the Internet. Recently, another famous game in Ukraine and in the world for the search of death has become known i.e. "Run or Die". According to the rules of the game the child should be in front of the car unexpectedly for its driver to cross the roadway. At the same time, the highest aerobatics is considered when a participant in a suicidal game during a run "touches" his body with the front of the car.

An example of similar games to death is "Red Owl", "Wake Me at 4:20", "Sea of Whales", "F57", "Quiet House", "Pink Fairies", "Dog Kite", etc. They, as in the "Blue Whale" and "Run or Die", have much in common. Investigators found that access to all these groups was strictly limited and was carried out at the discretion of the administrator of the "death group". Newly-made participants had to pass different tests and tasks, as the performance of which progressed the rating, which opens up access to new "opportunities": content and tasks related to the theme of death, suicide, self-injury and staging of these events. Death was propagated as the only correct and beautiful way out of difficult life situations. Cultivated depressive psychological states, the use of violence against themselves and others, and human values – family, friends, education – were criticized and laughed at.

If we talk about the legal assessment of the "games of death" data, then, of course, this is about crimes, not ordinary suicides. Indeed, according to the Department of Cyberpower Serhiy Demedyuk, "Suicide is a conscious self-deprivation of life. And here's a crime. They provoke: "Can you do it?" That's when they find out from the very large number of the weakest children i.e. those who cut their hands, show that they can do something for themselves – all the attention of the group is redirected to it. And the terrible countdown of

days starts, the "quest" that begins in the finale leads to the most terrible" [8].

Employees of the cyberpolice found out: often in "death groups", if the child refuses to kill himself, the criminals calculate the IP address of the participant and tell him that for this cowardice will have to answer his relatives. Moreover, according to the National Police of Ukraine, under the guidance of administrators for children, there are skilled psychologists who can quickly push them into suicide.

As we see, with the development of technical progress and evolution of humanity there are new, no less dangerous ways of committing with the help of social networks and the Internet chain of criminal acts that will encroach on human life and health. Unfortunately, neither Ukrainian society nor law enforcement agencies are ready for these new challenges.

#### IV. CURRENT LEGISLATION AND LAW ENFORCEMENT PRACTICE IN THE AREA OF RESPONSIBILITY FOR INCITEMENT TO SUICIDE USING SOCIAL NETWORKS AND THE INTERNET

On the one hand, it's terribly to state that no administrator of the "death group" of the social network has been adequately punished for committing suicides by their minor participants, but on the other hand it should be remembered: "There is no crime and punishment without a predetermined law" (Latin "Nullum crimen, nulla poena sine praevia lege poena lige"). After all, this principle of law is obliged to adhere to every state that considers itself to be legal and democratic.

Until the entry into force of March 7, 2018, the Law of Ukraine No. 2292-VIII "On Amendments to Article 120 of the Criminal Code of Ukraine on the Establishment of Criminal Responsibility for the Assistance to Suicide" [9], the actions of the administrators of "death groups" most closely fell within the indications of the crime "Incitement to Suicide", stipulated by Art. 120 of the Criminal Code of Ukraine. However, it is not true for 100%. The editorial of Part 1 of Art. 120 of the Criminal Code of Ukraine, which was in force until March 7, 2018 (before the entry into force of the above-mentioned law), provided the criminal liability for bringing another person to a suicide or suicide attempt, which, from the objective part of the crime, could only be committed in one of the following ways: 1) cruel treatment with him; 2) blackmail; 3) coercion for unlawful actions, or 4) the systematic humiliation of its human dignity. Qualifying signs of incitement to suicide or suicide attempt was recognized at the time of the commission specified in Part 1 of Art. 120 of the Criminal Code of Ukraine regarding a person who was materially or otherwise dependent on the perpetrator or on two or more persons (Part 2 of Article 120 of the Criminal Code of Ukraine) or if it was committed against a minor (Part 3 of Article 120 of the Criminal Code of Ukraine) [10].

Consequently, incitement to suicide, which was committed by March 7, 2018, without the use of coercion, blackmail, humiliation and the other stipulated in Art. 120 of the Criminal Code of Ukraine the ways (for example, by systematic inclining a person to a suicide) were not recognized as a criminal act. Therefore, law enforcement

agencies could not be prosecuted for committing suicide or attempted suicide, which was carried out through the use of social networks or the Internet. Furthermore, most suicide cases committed by adolescents who were members of so-called “death groups” did not have any criminal-law assessment at all. Therefore, the society faced an urgent question about how to solve this problem.

One of the possible ways of controlling children on the Internet was the introduction of a ban on registration in social networks for children under the age of 14 years to avoid child suicides. However, we believe that the use of this method of control will be inappropriate and illegal. First, such a restriction would not be in line with modern living conditions and will be, first of all, the violation of the child's right to access the Internet; and secondly, this ban will not prevent children from registering in social networks.

On February 8, 2018, the Parliament of Ukraine adopted the Law of Ukraine “On Amendments to Article 120 of the Criminal Code of Ukraine on the Establishment of Criminal Responsibility for the Assistance to Suicide, which was intended to establish criminal liability for any assistance to a person in committing suicide or attempted suicide [11].

To achieve this goal, this Law was laid down in Part 1 of Art. 120 “Conduct to suicide” of the Criminal Code of Ukraine in a new wording, establishing criminal responsibility for “proving a person to suicide or attempted suicide resulting from the cruel treatment, blackmail, systematic humiliation of his human dignity or systematic unlawful coercion of actions that contrary to her will, predilection for suicide, as well as other actions that contribute to committing suicide” [12]. As we have already noted, the new wording of this article came into force on March 7, 2018.

In our opinion, Art. 120 of the Criminal Code of Ukraine, as amended on February 8, 2018, will not facilitate the establishment of proper normative regulation of criminal responsibility for bringing a person to suicide or any assistance to a person in committing a suicide or attempting to commit the latter through the use of social networks or the Internet. We believe that the more successful option of teaching art. 120 of the Criminal Code of Ukraine would have been editorially containing a clear indication of the objective aspect of this criminal act, such as the use of the social network or the Internet. This would contribute to better practical application of this article, minimize mistakes when qualifying these criminal acts. After all, first of all, it is not clear from the contents of the current edition of Part 1 of Art. 120 of the Criminal Code of Ukraine that its norm covers cases of criminal liability for the proving or attempted suicide attempts under the influence of the Internet or social networks.

Secondly, we believe that incitement to suicide or attempted suicide is a computer crime, and therefore it would be much more effective if the parliament did not make changes to Part 1 of Art. 120 of the Criminal Code of Ukraine, and in this regard subjected the editors of the Criminal Code of Ukraine in part of the norms providing for criminal liability for computer crimes.

## V. WAYS TO IMPROVE THE LEGISLATION ON CRIMINAL RESPONSIBILITY FOR INCITEMENT TO SUICIDE WITH SOCIAL NETWORKS AND THE INTERNET

We believe that the acts of committing suicide with the use of social networks and the Internet should be classified as a category of computer crimes. Therefore, the norm, which would include criminal liability for the given act, should be found in Section XVI of the Criminal Code of Ukraine i.e. “Crimes in the field of the use of electronic computing machines (computers), systems and computer networks and telecommunication networks” [13].

The impassable placement of information in the networks that not only provokes children to deprive oneself of life, but also popularization of this information in social networks and the Internet is one of the most important causes of child suicide. Therefore, criminally punishable should be both the placement in the public access networks of suicide, and its popularization.

In Ukraine, we propose introducing criminal liability for the creation of “death groups” and games containing suicidal topics. It is necessary to foresee the criminal responsibility for inclining – blackmail, bribery, deception, persuasion, etc. – to children for suicide, as well as for the promotion and moral and psychological push of children to commit suicide. It is also necessary to criminalize the creation, use and distribution of botnets, that is, a network of computers infected with a malicious program that allows criminals to control computer devices remotely and use them to commit any illegal acts, including the distribution of content suicidal character.

We consider it necessary to establish a criminal responsibility for bringing not only to suicide, but also to commit self-harm.

## VI. CONCLUSIONS

Consequently, taking into account the global tendency to increase the crime rate on the Internet and the emergence of a new way of suicide, it is necessary to make timely changes to the criminal law that will meet the current challenges of the perpetrators. The current CC of Ukraine made an attempt to distinguish punishment for the criminalize acts of suicide or incitement to suicide by using social networks and the Internet. However, we believe that more effective and effective in counteracting and combating these crimes will be a tendency for them to be perceived not only as crimes against human life and health, but also, first and foremost, as computer crimes in legislation and law enforcement practice. Actually, it is with this in mind that the reform of criminal legislation and the practice of its application in Ukraine should be implemented.

## REFERENCES

- [1] Pojasnuval'na zapyska do proektu Zakonu «Pro vnesennja zmin do Kryminal'nogo kodeksu Ukrayny (shhodo vstanovlennja kryminal'noi' vidpovidal'nosti za spryjannja vchynennju samogubstva)» [Explanation for the Draft of the law “on changes of the Criminal Code of

- Ukraine on the criminal liability for incitement to suicide”] – [Electronic resource]. – Mode of access: [http://w1.c1.rada.gov.ua/pls/zweb2/webproc4\\_1?pf3511=58195](http://w1.c1.rada.gov.ua/pls/zweb2/webproc4_1?pf3511=58195). [in Ukrainian].*
- [2] Kryminal’nyj kodeks Ukrai’ny vid 5 kvitnja 2001 r. (v redakcii’ stanom na 18 kvitnja 2018 r.) [Criminal Code of Ukraine]. – [Electronic resource]. – Mode of access: <http://zakon2.rada.gov.ua/laws/show/2341-14>. [in Ukrainian].
- [3] Pojasnjuval’na zapyska do proektu Zakonu «Pro vnesennja zmin do Kryminal’nogo kodeksu Ukrai’ny (shhodo vstanovlennja kryminal’noi’ vidpovidal’nosti za spryjannja vchynennju samogubstva)» [Explanation for the Draft of the law “on changes of the Criminal Code of Ukraine on the criminal liability for incitement to suicide”] – [Electronic resource]. – Mode of access: [http://w1.c1.rada.gov.ua/pls/zweb2/webproc4\\_1?pf3511=58195](http://w1.c1.rada.gov.ua/pls/zweb2/webproc4_1?pf3511=58195). [in Ukrainian].
- [4] Pojasnjuval’na zapyska do proektu Zakonu «Pro vnesennja zmin do Kryminal’nogo kodeksu Ukrai’ny (shhodo vstanovlennja kryminal’noi’ vidpovidal’nosti za spryjannja vchynennju samogubstva)» [Explanation for the Draft of the law “on changes of the Criminal Code of Ukraine on the criminal liability for incitement to suicide”] – [Electronic resource]. – Mode of access: [http://w1.c1.rada.gov.ua/pls/zweb2/webproc4\\_1?pf3511=58195](http://w1.c1.rada.gov.ua/pls/zweb2/webproc4_1?pf3511=58195). [in Ukrainian].
- [5] Smertel’ni igry pidlitkiv: na zminu «Synim kytam» pryjshla sui’cydal’na «Chervona sova». [Death games of teens: instead of “Blue whales” appeared “Red owl”] – [Electronic resource]. – Mode of access: <http://expres.ua/news/2017/12/06/275162-smertelni-igry-pidlitkiv-zminu-synim-kytam-pryjshla-suyicydalna-chervona-sova>. [in Ukrainian].
- [6] «Grupy smerti» staly novym vyklykom dlja suspil’sta ta policii’. [“Groups of death” became a new challenge for the society and police] – [Electronic resource]. – Mode of access: <https://tsn.ua/ukrayina/sini-kiti-kiberpoliciya-viyavila-mayzhe-1000-grup-smerti-ta-poperedila-ponad-10-sprob-suyicidu-919927.html>. [in Ukrainian].
- [7] Dovedennja do samogubstva: pravovyj aspekt . [Incitement to the suicide: legal aspect] – [Electronic resource]. – Mode of access: <http://reporter.pl.ua/novini/sytuatsija/26463-dovedennja-do-samogubstva-pravovyj-aspekt>. [in Ukrainian].
- [8] «Grupy smerti»: golovnyj kiberpolicejs’kyj Ukrai’ny rozkazav podrobyci pro samogubstva pidlitkiv i skandal’nyh kytiv. [“Group of death”: the chief cyberpoliceman of Ukraine told about the details of teens’ suicides] – [Electronic resource]. – Mode of access: <https://nv.ua/ukr/publications/tse-vbivstva-na-jakih-zaroblajut-dorosli-golovnij-kiberpolitsejskij-ukrajini-rozkazav-pro-grupi-smerti-641854.html>. [in Ukrainian].
- [9] Zakon Ukrai’ny «Pro vnesennja zminy do stati 120 Kryminal’nogo kodeksu Ukrai’ny shhodo vstanovlennja kryminal’noi’ vidpovidal’nosti za spryjannja vchynennju samogubstva» № 2292-VIII vid 8 lютого 2018 r. [The law of Ukraine “On changes in article 120 of the Criminal Code of Ukraine on setting the criminal liability for incitement to suicide”] – [Electronic resource]. – Mode of access: <http://zakon2.rada.gov.ua/laws/show/2292-19/paran5#n5>. [in Ukrainian].
- [10] Kryminal’nyj kodeks Ukrai’ny vid 5 kvitnja 2001 r. (v redakcii’ stanom na 18 kvitnja 2018 r.) [Criminal Code of Ukraine]. – [Electronic resource]. – Mode of access: <http://zakon2.rada.gov.ua/laws/show/2341-14>. [in Ukrainian].
- [11] Zakon Ukrai’ny «Pro vnesennja zminy do stati 120 Kryminal’nogo kodeksu Ukrai’ny shhodo vstanovlennja kryminal’noi’ vidpovidal’nosti za spryjannja vchynennju samogubstva» № 2292-VIII vid 8 lютого 2018 r. [The law of Ukraine “On changes in article 120 of the Criminal Code of Ukraine on setting the criminal liability for incitement to suicide”] – [Electronic resource]. – Mode of access: <http://zakon2.rada.gov.ua/laws/show/2292-19/paran5#n5>. [in Ukrainian].
- [12] Kryminal’nyj kodeks Ukrai’ny vid 5 kvitnja 2001 r. (v redakcii’ stanom na 18 kvitnja 2018 r.) [Criminal Code of Ukraine]. – [Electronic resource]. – Mode of access: <http://zakon2.rada.gov.ua/laws/show/2341-14>. [in Ukrainian].
- [13] Kryminal’nyj kodeks Ukrai’ny vid 5 kvitnja 2001 r. (v redakcii’ stanom na 18 kvitnja 2018 r.) [Criminal Code of Ukraine]. – [Electronic resource]. – Mode of access: <http://zakon2.rada.gov.ua/laws/show/2341-14>. [in Ukrainian].

# The Impact of GDPR on IT/IS

Vlasta Svatá

Department of Systems Analysis, University of Economics, Prague, W. Churchill sq. 3, 130 00 Prague 3, email:svata@vse.cz

**Abstract:** The article focuses on GDPR and emphasizes the implications of ensuring compliance in IS/IT. It stresses the importance of information security management and data governance. The main output is the table outlining the main responsibilities of data controllers and processors and their impact on IS / IT management.

**Keywords:** GDPR, information security, data governance, IT/IS perspectives of GDPR.

## I. INTRODUCTION

IT professionals can act as strategic partners to businesses currently working toward compliance with the European Union General Data Protection Regulation (GDPR), scheduled to come into enforcement on 25 May 2018. Even to the fact, that GDPR compliance is not solely a technology issue (it requires attention and remediation expertise from various functions within the business, e.g. human resources, legal, compliance, marketing, communications) technology acts as a common denominator across business processes and plays a significant role in the collection, processing, storage and transfer of personal data [3]. The main aim of the article is to clarify the status of personal data protection in information security management, and thereby partially reduce concerns about the introduction of regulation into practice. At the same time, some new aspects that the new regulation brings to IT management should be highlighted.

## II. PROTECTION OF PERSONAL DATA PRIVACY, INFORMATION SECURITY AND DATA GOVERNANCE

Confidentiality / privacy is a fundamental right, necessary for autonomy and protection of human dignity, serving as the basis on which other human rights are built. On the other hand IT fundamentally restricts the right to privacy being integrated, globalized and mobile. This fact is counterbalanced by the constant improvement of frameworks for the introduction of controls into the IS / IT environment. All of them should artificially balance the loss of privacy. The development of such countermeasures range from securing the primitive need to "be alone - give me a peace" up to complex concepts (legal, socio-psychological, economic or political). Some of the concepts are more reactive (GDPR is an example) some are more proactive, e.g. European Data Protection: Coming of Age ). This document takes in account seven categories of privacy:

- Privacy of the person: protection against unauthorized body failure (genetic tests, blood tests, implants, ...).
- Privacy of behavior and actions: protection of ideas, emotions, orientation - sensitive information (camera systems, police body cameras).

- Privacy of communication: protection of all communication channels (printed, voice, video, digital) (hidden microphones, mail, postal services).
- Privacy of association: the right to associate with other persons without unauthorized monitoring and marginalization (DNA tests for ethical proof, employee release based on DNA tests, ...).
- Privacy of data and image (information): protection of personal information in all forms (leakage of financial, health information, dissemination of images without the knowledge of persons).
- Privacy of thoughts and feelings: protection against their spread or use against persons (requesting password for access to the social network when recruiting, requesting information on religion and political orientation).
- Privacy of location and space (territorial): protection against technology that can monitor the location, space, and general environment of an individual (video, drones, work and home monitoring).

|                                 | Social media | Cloud computing | Apps | Big Data Analytics | Internet of Things | BYOD | Tracking and surveillance |
|---------------------------------|--------------|-----------------|------|--------------------|--------------------|------|---------------------------|
| Privacy of the person           |              |                 | x    | x                  | x                  |      | x                         |
| Privacy of behaviour and action | x            |                 | x    | x                  | x                  | x    | x                         |
| Privacy of communication        | x            | x               | x    | x                  | x                  | x    |                           |
| Privacy of data and image       | x            | x               | x    | x                  | x                  | x    | x                         |
| Privacy of thought and feelings | x            | x               | x    | x                  | x                  | x    | x                         |
| Privacy of location and space   | x            | x               | x    | x                  | x                  | x    | x                         |
| Privacy of association          | x            | x               | x    | x                  | x                  |      | x                         |

Fig. 1. Privacy categories and technologies [2].

Fig. 1 shows how separated IT influence the privacy categories.

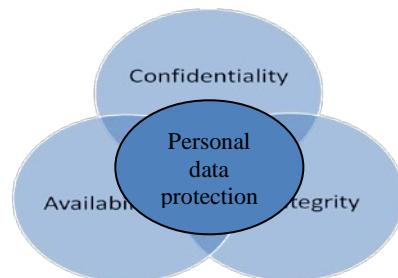


Fig. 2. Information security CIA Triad.

Great majority of international reactive frameworks for information security are designed to protect the confidentiality, integrity and availability of computer system data from those with malicious intentions. Confidentiality, integrity and availability are sometimes referred to as the CIA Triad of information security. Based on this definition we can conclude, that protection of personal data is preservation of confidentiality, integrity and availability of data relating to a living individual who is or can be identified either from the data or from the data in conjunction with other information. Protection of personal data thus can be viewed as application of information security controls over the specific types of data – personal data. The base for GDPR is thus the compliance with information security frameworks, but it is only a necessary but not a sufficient condition. GDPR goes beyond the information security frameworks, as its main aim is to improve data governance.

Data governance is a term used to describe the overall, comprehensive process for controlling not only the data security (CIA Triad) but the efficiency, effectiveness, relevance and compliance as well. Data governance consists of the processes, methods, tools, and techniques to ensure that data is of high quality, reliable, and unique (not duplicated), so that downstream uses in reports and databases are more trusted and accurate. A data governance program should be a part of an IT governance program. While data/information security is mainly in responsibility of IT professionals (processors or third parties), data governance is in responsibility of business managers – controllers that determine the purposes and means of the processing of personal data. GDPR compliance is thus the corporate responsibility of the data controller, not of the DPO, internal auditor or CIO.

### III. PERSONAL DATA AND PERSONAL DATA PROCESSING

Before we will discuss the impact of GDPR on IT management it is necessary to specify who does the GDPR apply to and what data does the GDPR apply to.

The GDPR applies to processing carried out by organizations operating within the EU. It also applies to organizations outside the EU that offer goods or services to individuals in the EU. Transfers of personal data outside the EEA (European Economic Area) are not allowed unless the country has an adequate level of protection for the processing of personal data. For instance in 2016 the European Commission approved the Privacy Shield, enabling easy transfer of personal data from the EU to selected companies without the need to obtain permission from the national data protection authority or the conclusion of a standard contractual clause with the US data processor. The GDPR applies to both the ‘controllers’ and ‘processors’. The controllers determine the purposes and means of processing personal data and ensure that the contracts with processors comply with the GDPR. They are not relieved of their obligations where a processor is involved and they shall be responsible for, and be able to demonstrate, compliance with the GDPR principles. The processors are responsible for processing personal data on behalf of a controller. They are required to maintain records of personal data and processing

activities and they will have legal liability if they are responsible for a breach. The GDPR does not apply to certain activities including:

- covered by the Law Enforcement Directive
- processing for national security purposes
- processing carried out by individuals purely for personal/household activities.

The GDPR applies to ‘personal data’ (both automated and manual filing systems) meaning any information relating to an identifiable person who can be directly or indirectly identified in particular by reference to an identifier. In case of manual data it must be accessible according to specific criteria. This could include e.g. chronologically ordered sets of manual records containing personal data. Special categories of personal data are sensitive data, pseudonymised data and children’s data. Sensitive data include genetic data, and biometric data where processed to uniquely identify an individual. Pseudonymised data (coded data) can fall within the scope of the GDPR depending on how difficult it is to attribute the pseudonym to a particular individual. Examples are all types of electronic traces, e.g. Internet proxy addresses and cookie identifiers. Children’s data processing needs the consent of the holder of parental responsibility and the data need specific protection as children may be less aware of the risks. Personal data that are excluded from compliance are:

- data relating to criminal convictions and offences
- organizational data
- data of deceased persons
- data to help prevent crime (investigation, detection, prosecution)
- data not arranged according to the specified points of view
- anonymous data (statistics, research).

Personal data processing is any act or set of acts that the controller or processor systematically performs with personal data by automated or other means. Examples are collecting, recording, arranging, structuring, storing information, accessing, editing or editing. This definition of data processing is no surprise, but the Article 5 of the GDPR requires that personal data shall be

- processed lawfully, fairly and in a transparent manner in relation to individuals;
- collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes.

The requirement to have a lawful basis in order to process personal data is not new. It replaces and mirrors the previous requirement to satisfy one of the ‘conditions for processing’ under the Data Protection Act 1998. However, the GDPR places more emphasis on being accountable for and transparent about the lawful basis for processing [1]. The identification of the lawful basis is important not only because the organizations should provide people with information about their lawful basis for processing (this information must be covered in the privacy notice) but it has a big consequences on to IT/IS area. The reason is, that the lawful basis for processing can also affect the rights that are to be available to individuals.

There exist the six lawful bases:

- Consent - should be given by a clear affirmative act of the data subject's agreement to the processing of personal data relating to him or her, such as by a written statement, including by electronic means, or an oral statement.
- Contract - processing should be lawful where it is necessary in the context of a contract or the intention to enter into a contract.
- Legal obligation - legitimate basis, laid down by law including the necessity for the performance of a contract to which the data subject is party.
- Vital interest - it is necessary to protect an interest which is essential for the life of the data subject or that of another natural person (e.g. humanitarian purposes).
- Public task - processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority; the processing should have a basis in Union or Member State law.
- Legitimate interest - e.g. where there is a relevant relationship between the data subject and the controller in situations such as where the data subject is a client or in the service of the controller, when it is necessary for the purposes of preventing fraud and for direct marketing purposes.

The GDPR provides the following rights for individuals:

- The right to be informed
- The right of access
- The right to rectification
- The right to erasure
- The right to restrict processing
- The right to data portability
- The right to object
- Rights in relation to automated decision making and profiling.

Fig. 3 shows the relevance between the six lawful bases and voluntary rights for individuals (the others are obligatory in each case).

#### IV. IMPORTANT ISSUES FROM IT/IS PERSPECTIVE

Despite the fact, that GDPR being data governance regulation, not the data security regulation, is in accountability of board and business executives, IT professional both from the controller's and processor's organizations are to be expected to address a range of tasks

|                     | Right to erasure | Right to portability | Right to object |
|---------------------|------------------|----------------------|-----------------|
| Consent             |                  |                      | x               |
| Contract            |                  |                      | x               |
| Legal obligation    | x                | x                    | x               |
| Vital interest      |                  | x                    | x               |
| Public task         | x                | x                    |                 |
| Legitimate interest |                  | x                    |                 |

Fig. 3. Relationships between the lawful bases and rights [1].

that goes beyond their existing responsibilities. The extent of these tasks is influenced by the following three aspects that are specific for each organization:

- Decision about the lawful base for personnel data processing – has an impact on the scope of GDPR application (see Chapter III)
- Codes of conduct – guidance on the implementation of appropriate measures and on the demonstration of compliance by the controller or the processor. They should include identification of the risk related to the processing, their assessment in terms of origin, nature, likelihood and severity, and the identification of best practices to mitigate the risk.
- Designation of the data protection officer – represents the new role within the organization and thus needs new redefinition of the RACI chart and redesign of the IT and business processes.

Next table provides the summarization of the controller's responsibilities and their impact on IT/IS.

TABLE 1. OVERVIEW OF CONTROLLERS AND PROCESSORS  
GDPR RESPONSIBILITIES AND THEIR IMPACT ON IS / IT MANAGEMENT

| Responsibilities of the controller and processor  | Impact on IT/IS   |
|---|---|
| Data protection principles:<br><ul style="list-style-type: none"> <li>• lawful, fair and transparent processing</li> <li>• collected for specified, explicit and legitimate purposes</li> <li>• minimization</li> <li>• accurate data</li> <li>• kept no longer than is necessary</li> <li>• secure data</li> </ul> | Provide consultancy what, where and how long are personal data collected, processed and stored  |
| Implementation of the appropriate technical and organizational measures   | Implement methods for the pseudonymisation and encryption of personal data<br><br>Ensure information security (confidentiality, integrity, availability) by choosing and implementation of the appropriate framework (ISO 27000, Cobit 5, Cobit Security Baseline, ..); impact on: <ul style="list-style-type: none"> <li>• business continuity (back up and disaster recovery plans)</li> <li>• risk assessment</li> <li>• access controls</li> <li>• physical security</li> </ul> Do not engage another processor without prior specific or general written authorization of the controller<br><br>Check all contracts that are binding on the processor whether they sets out the subject-matter and duration of the processing, the nature and purpose of the processing, the type of personal data and categories of data subjects and the |

|  |  |
|--|--|
|  | <p>obligations and rights of the controller.</p> <p>Continuous testing, assessing and evaluating the effectiveness of technical and organizational measures</p>  |
| Approval of certification mechanism (is voluntary)                           | Adherence to the approved certification mechanism and cooperation with the certified bodies  |
| Notification of a personal data breach to the supervisory authority          | Immediate notification the controller of a personal data breach in a formal way (breach nature, consequences, taken measures, etc.)  |
| Data protection impact assessment  | Assessment of the impact of new IT on the personal data protection (cloud computing, big data, IoT, BYOD, ...)   |
| Designation of the data protection officer (DPO)                             | <p>Support the DPO in performing the tasks (access to personal data and processing operations)</p> <p>Ensure that DPO tasks and duties do not result in a conflict of interests</p> <p>Provide IT/IS consultancy</p>   |
| Codes of conduct   | <p>Adherence to the codes of conduct; impact on:</p> <ul style="list-style-type: none"> <li>• declaration, that personal data processing is fair and transparent</li> <li>• the pseudonymisation of personal data (e.g. whether System design permits the attribution of pseudonymized data to natural persons, domain segregation is applied to separate attribution data from pseudonymized data; and access to meta-data is appropriately restricted)</li> <li>• the information provided to the public and to data subjects</li> <li>• the notification of personal data breaches</li> <li>• the transfer of personal data to third countries</li> <li>• the information security measures and procedures</li> </ul> <p>Cooperation with the accredited body while monitoring compliance</p> |
| Transfers of personal data to third countries or international organizations | <p>Provide consultancy as regard the appropriate safeguards, and condition that enforceable data subject rights and effective legal remedies for data subjects are available</p> <p>Provide contractual clauses about safeguards</p>   |
| Lawfulness of processing   | <p>Consent processing</p> <ul style="list-style-type: none"> <li>• the need to record the consents, purpose and validity and to check them in personal data processing</li> <li>• to be able to withdraw the consent at any time</li> </ul>  |
| Rights of the data subject   | <p>To check possibilities how to automate the realization of the separate rights; impact on</p> <ul style="list-style-type: none"> <li>• authentication of the data subject enforcing its law</li> <li>• personal data encryption</li> </ul>   |

|                                  |  |
|----------------------------------|--|
|                                  | <ul style="list-style-type: none"> <li>• changes in process, data, application models (additional controls, identifiers, functions)</li> <li>• changes in application interfaces (menus) supporting the communication with data subject</li> <li>• process analysis for separate rights enforcement</li> </ul>   |
| Records of processing activities | <p>Provide consultancy about the items needed, mainly:</p> <ul style="list-style-type: none"> <li>• the purposes of the processing</li> <li>• the categories of data subjects and of the categories of personal data</li> <li>• the categories of recipients</li> <li>• transfers of personal data to a third country</li> <li>• time limits for erasure</li> <li>• general description of the technical and organizational security measures</li> <li>• the name and contact details of the processor(s)</li> </ul> |

## V. CONCLUSION

IT plays a dual role in the protection of personal data: in one it poses a threat, the other is an effective protection tool. Balancing these two roles is an endless and costly task for all organizations. As a consequence GDPR cannot be viewed as a sprint to finish line. It represents one of the great opportunities that provides the basis for deepening collaboration between business executives and IT professionals. In many cases, IT professionals can assure managers that the required controls are already implemented or can be done automatically, in other situations they can point out the IT risks that they pose to the protection of personal data. In any case, without deeper ongoing cooperation and communication between these parties, ensuring compliance with the GDPR will be the only investment without any value for business.

## REFERENCES

- [1] Information Comissionair's Office, Guide to the General Data Protection Regulation, 22 March 2018
- [2] ISACA Privacy principles and program management Guide, <http://www.isaca.org/Knowledge-Center/Research/ResearchDeliverables/Pages/ISACA-Privacy-Principles-and-Program-Management-Guide.aspx>
- [3] O.Osagiede Beyond GDPR Compliance – How IT Audit Can Move from Watchdog to Strategic Partner, isaca.org
- [4] Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (GDPR)

# Smart Grid

# The Role of Smart Meters in P2P Energy Trading in the Low Voltage Grid

Christina Sigl<sup>1</sup>, Alexander Faschingbauer<sup>1</sup>, Andreas Berl<sup>1</sup>,  
 Jakub Geyer<sup>2</sup>, Rudolf Vohnout<sup>2</sup>, Miloš Prokýšek<sup>2</sup>

1. Institute for Applied Informatics, Deggendorf Institute of Technology, Freyung, Germany,  
 email: {christina.sigl, alexander.faschingbauer, andreas.berl}@th-deg.de

2. Institute of Applied Informatics, University of South Bohemia, Ceske Budejovice, Czech Republic,  
 email: {geyer, vohnout, prokysek}@prf.jcu.cz

**Abstract:** There are different approaches for energy trading shown in the paper. The peer-to-peer approach is the most interesting one of them. Therefore, the underlying subsystem is essential for fast and accurate trading. Smart meters are one of the most important parts within the infrastructure to provide high quality information and services for smart contracting and also for controlling the grid. Therefore, smart meters have to fulfill dedicated requirements. Especially the lowest layer of smart metering – the smart meter itself – is considered here. A very basic view on data acquisition and a possible architecture for a test bench are presented in this paper.

**Keywords:** smart grid, P2P, energy, trading, smart meter.

## I. INTRODUCTION

Smart grids are based on smart measurement and control of energy supply, transportation and demand. The intelligent control of all components can be seen as a virtual grid, where energy can be traded over short (low voltage grid connected to a sub-station) and/or long distances (middle- and high voltage grid). In both cases the complexity of the control and trading system is very high.

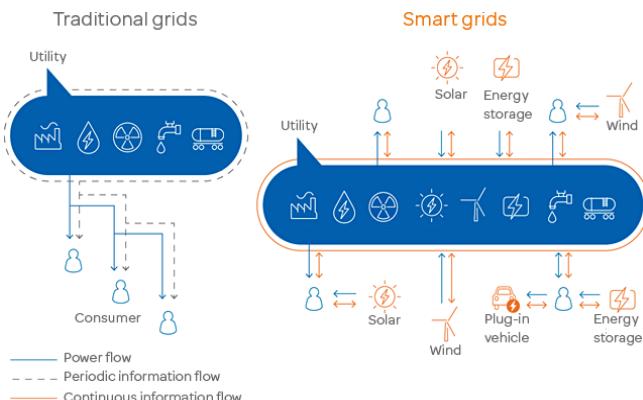


Fig. 1. Grid schematic [1].

Fig. 1 compares the traditional grid with a smart grid. In traditional grid architectures there are producers (utility) and consumers. Whereas in the smart grid renewable energy sources and storages are integrated and therefore a consumer can take the role of a producer as well.

In case of peer-to-peer (P2P) trading also both roles (prosumer) are used by each participant, which changes their

behavior. As such, to enable this system to be dynamic and flexible it needs respective agile components to be able to provide production-consumption control. Thus, smart meters (SM) are important in such a P2P network. SMs provide information about power consumption and distribution for billing (longer time intervals as for grid monitoring). Maybe, they can be used to detect problems in the power grid, too [1]. Within a P2P network SM are very important. Therefore, required technologies, protocols and data quality must be analyzed regarding to P2P trading. Thus, a simplified consideration of measurement accuracy is given and some gaps in the measurement chain (e.g. timing issues) are depicted. Additionally, a test bed, especially for analyzing SMs acquisition accuracy, is introduced. Moreover, in P2P trading, business models and means of trustworthy evidence of the contract (for example smart contracting/block chain) must exist. Hence, P2P systems, business models as well as energy pricing models are considered.

The Paper is organized as follows. In section II, related work is given. Section III presents P2P energy trading models and a comparison of energy prices between Germany and Czech Republic. Information, that SMs have to provide for P2P trading, are shown in section IV. Required technologies and protocols are pointed out in section V. The significance of data quality, for P2P trading, is stated in section VI. A test bed setup for analyzing SMs is shown in section VII. Finally, the paper is concluded in section VIII.

## II. STATE OF THE ART

P2P trading, smart contracting and micropayment are modern concepts, which are widely used nowadays. There are several pilot projects and examples dealing with this topic. Murkin et al. show an example platform for P2P energy trading using the block chain technology. This makes the trading process accurate, authenticated and secure [2]. Alvaro-Hermana et al. use electric vehicles for P2P energy trading [3]. In general, there are different approaches of P2P energy trading systems on that we will take a closer look in this paper.

Matamoros et al. investigated P2P trading between two micro grids, thus looking at central versus distributed control [4]. Zhang et al. submit that communication and control networks are very important for P2P energy trading. They also show a future scenario of P2P energy trading [5]. For

control of the P2P trading system data from smart metering must be used.

P2P energy trading bases on timing, a reliable communication layer and accurate metering. Marshall et al. did some investigations/simulations about the impact of accurate metering. Most commercial metering systems measure net flow only in intervals of 30 seconds or 30 minutes, which is a barrier for accurate accounting. Therefore, sub-second level timing is required. Economic impacts through faster energy fluctuations, than the measurement intervals, lead to economic inefficiencies and also possible inaccuracies through meter timing (the question is stated, which time period is the best) [6]. Capodieci et al. [7] present a hardware/software solution for energy trading using agents which use only six time intervals per day for trading. The hardware architecture consists of a real SM which is connected to a SM gateway through a SM interface. The data is sent to an energy trading platform.

Nonetheless, SM accuracy also depends on temperature effects of SMs [8]. SM accuracy is determined considering ADC resolution, SINAD and THD [9]. Thus, we take a closer look on data quality that is required especially for P2P trading and to evaluate SMs a test environment is set up.

### III. P2P ENERGY TRADING

In general, there are differences concerning the electricity market of the neighboring countries Czech Republic (CZ) and Germany. In CZ, the current electricity market can be divided into several levels or areas. At first, there is a market for trading energy between producers and suppliers operated by market operator (OTE [10]). There also exists Power Exchange Central Europe (PXE [11]). This market is powered by EEX and provides also services for end-users, especially bigger consumers as municipalities or SMEs.

In Germany, there are two different business models for the electricity market. The first one is the traditional producer/consumer model. The second one can be called prosumer model. It is based on own production and consumption. Every customer (private/company) is connected to the grid by a distribution network operator (DSO) (e.g. Bayernwerk AG). The DSO is a direct customer of the four transmission system operators (TSO) (e.g. TENNET). Each customer has a contract with an electricity supply company (e.g. E.ON), which does billing. Electric energy is typically traded on the stock market.

In a smart grid environment, the control of consumer behavior and demand could be based on many things like social influence or responsibility, but the price for energy is the most important factor. Hence, the metering of consumption and production is the crucial part of all P2P trading systems (See TABLE 1 for overview).

#### *Current pricing model*

Electricity for smaller consumers is in CZ delivered by energy suppliers, usually based on an end-user agreement. For households, there are two common pricing models. For smaller consumers (mainly in high density areas with prefabricated houses) the price of energy is constant and calculated from several parts, mainly from the price of power,

distribution costs and taxes. For bigger consumers like houses there is a system of high and low tariffs.

Fig. 2 shows pricing components of consumed energy (E.ON, distribution rate D01d, 2018, [12], for Germany see [13]) excluding fix costs (monthly fee for consumption point, reserved input fee).

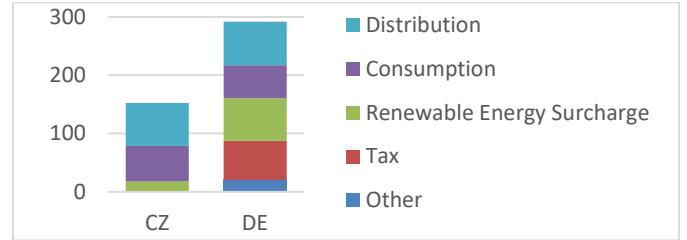


Fig. 2. Energy price components (excluding fix costs).

#### *Scheduled flexible pricing*

A flexible pricing model is based on PXE trading system and uses intraday trades. The system is now usually operated manually or with a low level of automation. This kind of trading system is a good base for further P2P business models and more advanced solutions. The amount of traded energy is based on prediction of consumption and could be supported by data gathering from smart buildings. For trading in PXE a license is required. Therefore, a mediator with license (usually an energy supplier) is an easier choice for end-users.



Fig. 3. Energy source matching [14].

#### *Linking energy production resources*

The system of linking/matching produced energy with consumers is based on flexible pricing and fluctuating power production of renewable energy sources. The idea is to offer cheaper energy to consumers when there is a surplus of energy and provide clear data about how the energy is produced. Consumers can then prioritize from which energy source and for what price they want to buy energy and move some energy consumption tasks accordingly. This allows savings for consumers while it also helps balancing local energy production and consumption.

#### *P2P energy market*

P2P energy market is based on a similar idea as the linking of energy production resources, the idea of balancing local energy production and consumption. However, in this case, any consumer can also become producer (prosumer) and his energy surplus is primarily offered to other local consumers (usually for a better price). If the offer is accepted by another local consumer, the transaction is realized (often using block chain mechanisms). Any additional surplus (or if there is no consumer at the time) is bought by a distribution company according to the agreed pricelist.

TABLE 1. P2P ENERGY TRADING PRINCIPLES OVERVIEW

|                                      | <b>Scheduled flexible pricing</b>   | <b>Linking energy production resources</b>  | <b>P2P Energy market</b>  | <b>P2P Energy trading platform (with battery storages)</b>  |
|--------------------------------------|---|---|---|---|
| <b>Companies / Projects</b>          | PRE   | Piclo, Vandebron, AmperMarket   | TransActive Grid, PeerEnergyCloud   | SonnenCommunity, Lichtblick Swarm Energy  |
| <b>Objectives</b>                    | Dynamic pricing based on estimated electricity production   | Linking electricity demand and local energy resources   | Direct local energy trading   | Distributed energy trading with power reserves and grid balancing capabilities  |
| <b>Peers</b>                         | Producers-Distributors-Consumers  | Producers-(Distributors)-Consumers  | Prosumers – Prosumers   | Prosumers – Prosumers   |
| <b>Key Features</b>                  | <ul style="list-style-type: none"> <li>• Price driven energy consumption estimates</li> <li>• Weather prediction</li> <li>• Consumption prediction</li> </ul> | <ul style="list-style-type: none"> <li>• Local energy production profiles</li> <li>• User consumption visualization</li> <li>• User energy resources preferences</li> </ul>   | <ul style="list-style-type: none"> <li>• Tokenization</li> <li>• P2P payments (block chain)</li> </ul>  | <ul style="list-style-type: none"> <li>• P2P payments (block chain)</li> <li>• User consumption prediction</li> <li>• Weather prediction</li> <li>• Virtual grid</li> </ul>   |
| <b>Infrastructure Level</b>          | Any   | Micro-grids / grid-cells  | Micro-grids / grid-cells  | Any   |
| <b>Smart Meter / Gateway Demands</b> | <ul style="list-style-type: none"> <li>• Daily / weekly / monthly readings</li> </ul>   | <ul style="list-style-type: none"> <li>• Daily / weekly / monthly readings</li> </ul>   | <ul style="list-style-type: none"> <li>• Readings several times per hour</li> <li>• P2P market support (online communication)</li> </ul>  | <ul style="list-style-type: none"> <li>• Readings several times per hour</li> <li>• P2P market support (online communication)</li> <li>• Gathering user data consumption (profile)</li> </ul>   |
| <b>Prosumers Control</b>             | Manual / parametric consumption adjustment (scheduled)  | Manual / parametric prosumers adjustment (scheduled)  | Manual / parametric prosumers adjustment (dynamic)  | Dynamic control based on user profile, weather prediction and grid (community) demands  |
| <b>Benefits</b>                      | <ul style="list-style-type: none"> <li>• Dynamic pricing for consumers</li> <li>• Load distribution more optimized to production</li> </ul>                   | <ul style="list-style-type: none"> <li>• Local production and distribution more optimized to production</li> <li>• More transparent billing information</li> <li>• Price reduction for adaptive consumers</li> <li>• Possible direct support of renewable energy sources</li> </ul> | <ul style="list-style-type: none"> <li>• Local consumption and distribution more optimized to production</li> <li>• Price reduction for consumers buying from prosumers</li> <li>• Higher selling price for prosumers</li> </ul>  | <ul style="list-style-type: none"> <li>• Consumption and distribution more optimized to production</li> <li>• Price reduction for consumers buying from prosumers</li> <li>• Higher selling price for prosumers</li> <li>• Peaks shaving</li> <li>• Power reserves</li> <li>• Grid balancing</li> </ul> |
| <b>URLs</b>                          | <ul style="list-style-type: none"> <li>• <a href="https://www.pre.cz">https://www.pre.cz</a></li> </ul>   | <ul style="list-style-type: none"> <li>• <a href="https://piclo.uk/">https://piclo.uk/</a></li> <li>• <a href="https://vandebron.nl">https://vandebron.nl</a></li> <li>• <a href="http://www.ampermarket.cz/">http://www.ampermarket.cz/</a></li> </ul>                             | <ul style="list-style-type: none"> <li>• <a href="https://lo3energy.com/">https://lo3energy.com/</a></li> <li>• <a href="https://TransActive.Grid">TransActive Grid)</a></li> <li>• <a href="http://www.peerenergycloud.de/">http://www.peerenergycloud.de/</a></li> <li>• <a href="http://www.smartmpower.com/">http://www.smartmpower.com/</a></li> </ul> | <ul style="list-style-type: none"> <li>• <a href="https://sonnenbatterie.de/en/sonnenbatterie">https://sonnenbatterie.de/en/sonnenbatterie</a></li> <li>• <a href="https://www.lichtblick.de/">https://www.lichtblick.de/</a></li> </ul>  |

### P2P energy trading platform (with battery storages)

A P2P energy trading platform consist of 3 key parts:

- P2P energy market
- Virtual grid with distributed battery storages
- Grid operator/distributor cooperation

Surplus energy can be stored and used in moments of low production (peak shaving) if a battery storage is added to the prosumer system. Based on the energy consumption profile (done by prediction), the stored energy can also be sold. With grid operator/distributor participation this can help balancing even across multiple grid cells.

### IV. SMART METERS IN P2P TRADING

Especially in the P2P trading market a smart metering solution is required, which is able to actively participate in the smart shared grid infrastructure. This system has to report electricity consumption values and trends. Additionally, it has to provide information for P2P trading to enable it. Such information is:

- Current demand of power consumption by a household/SME
- Total power budget available for trading in the (micro) grid shared (sub-) infrastructure
- Tariff and/or price per kWh of redundant power to be used
- Price for using/renting the infrastructure (distribution costs) for such model
- Length of the contract (for example one hour) and smart contract block chain evidence
- Amount of electricity units to be contracted
- Guarantee power supply to be provided for at least contacted time frame and for unit price agreed

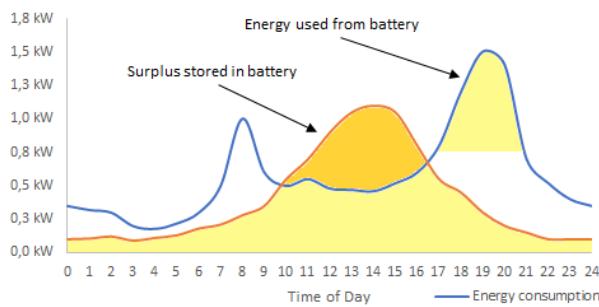


Fig. 4. Prossumer energy consumption profile (with photovoltaics and battery).

The role of SMs in such a situation is to provide control over the implementation phase of the P2P (smart) contract. As such, the internal real time clock accuracy and synchronization should be not lower than  $5 \times 10^{-3}$  seconds. Because of the nature of smart contracts, SMs used in the smart contract enabled grid infrastructure (with tariff less trading) have to be able to store enough historical values. Therefore, SMs primary oriented on distributed infrastructures and based on tariff templates will not be suitable for this scenario. Smart metering solutions available on the market differ in several aspects. For our model, the minimum accuracy level (we count only on transformer connected SMs) must be 0.5 S (better 0.2 S) to be able to provide a fair smart contract for parties involved even in low power demand conditions.

### V. TECHNOLOGY/PROTOCOLS

From the standardization point of view, SMs can be divided into two groups:

- Using proprietary communication protocols and security. Backward compatible with SM standard protocols such as DLMS or PRIME Alliance.
- Using standardized metering devices, certified by DLMS and/or PRIME Alliance.

The first group is mainly intended for island or isolated smart grid installations, where no interactions with other smart grid domains are expected and usually leads to vendor lock-in situations. Most of the smart grid solutions use standardized (certified) interoperable equipment to be able to exchange information not only within the same grid (domain), but also outside, to other installations supporting the same communication languages. For P2P trading we need only selected elements of xDLMS protocol structure protocol structure, avoiding those intend for energy distributor purposes.

### VI. DATA QUALITY

Data quality is one of the most important topics in the context of P2P energy trading. All parts of the described subsystem can affect data quality. This starts at the very low layer, where the signals are tapped from the grid lines, as it can be seen in the upper part of Fig. 5. Also the timing of the whole system must be known. Some blocks have a non-deterministic timing, which can lead to problems if the fluctuation rate of supply and demand is higher than the data processing by the P2P trading layer.

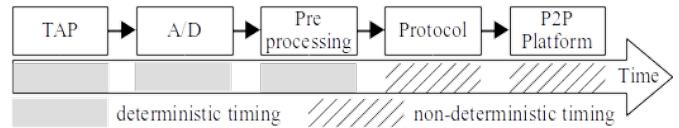


Fig. 5. Timing of Data Acquisition Process

The accuracy of the measured values is another important topic. There are several parameters which affect accuracy and data quality:

- Sensor type, attached to the grid lines
- Temperature
- Data reduction
- Communication protocols/gateways
- Processing speed
- Acquisition bandwidth (harmonics)

Fig. 6 shows a simulation of the measurement error, resulting from lowering the acquisition bandwidth, especially for reactive power in dependence of the number of harmonics considered.

The following assumptions were done: Line voltage  $U_{rms} = 230$  V, line current  $I_{rms} = 10$  A, grid frequency  $f = 50$  Hz, sample rate  $F_s = 100$  kHz. The signal of the line current is a sine wave superimposed with harmonics (damping by 100).

The calculation is done 100 times, starting with adding 100 harmonics to the base signal and decrementing it after each calculation step. For each loop the relative error is calculated, using the actual test signal in relation to the first signal.

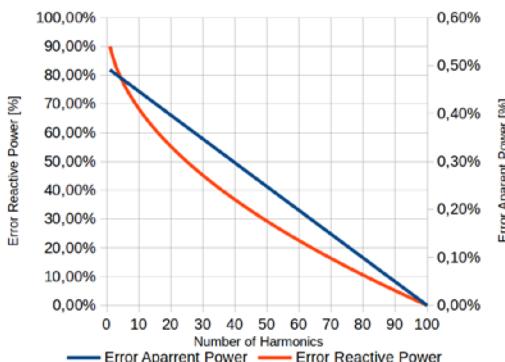


Fig. 6. Measurement Error in Dependence of Considered Harmonics

## VII. SMART METER TEST SETUP

The whole data processing chain from grid to the control system must be well known for well-balanced P2P trading. Fig. 7 depicts a simple test setup for testing SMs concerning their time delay.

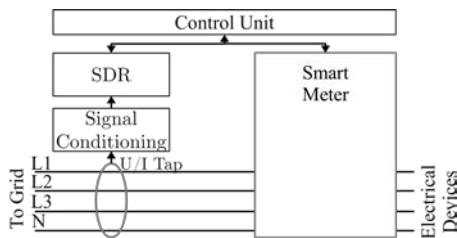


Fig. 7. Architecture of a SM test setup.

The grid signals are fed into a Software Defined Radio (SDR) through a voltage/current tap. The type of the current tap is a Hall sensor with higher frequency bandwidth compared to current transformers. Both signals (voltage and current) are applied to the SDR through filters, amplifiers, attenuators and impedance matching circuitry. The SDR is directly connected to a computer for controlling and processing the incoming data. The used SDR provides extremely high flexibility through the possibility of reconfiguration. The SM is also connected to the grid lines and considered as a black box, because the signal conditioning and processing is not accessible from the outside. A digital interface allows only access on the measurement results. Both acquisition systems are connected to the same computer to get synchronized datasets, for easier post processing and comparison.

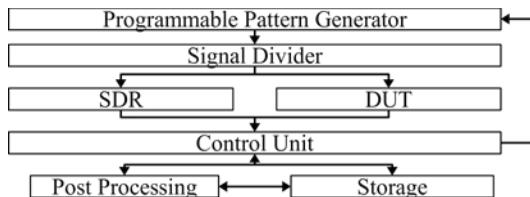


Fig. 8. Test Setup.

Fig. 8 shows the test setup in detail. The device under test (DUT) is a SM. DUT and SDR are both connected to a programmable pattern generator. Test signals, like a sine wave with a frequency of 50 Hz, superimposed by a variation of distortion signals of different speed, are applied. The distortion signals frequency is varied until the SM can't detect any further changes.

## VIII. CONCLUSION

This paper is focused on differences between traditional and smart grids from the view of business and metering. We identified the important role of smart metering and its necessary parameters for different business models of a P2P trading systems. We also discussed the price of energy for end-users/consumers on Czech and German energy market, where the component of energy distribution was identified as the most important and its reduction due the innovative trading scenarios could be the crucial motivation factor for P2P trading spread. Finally, we introduced a schema of a smart metering test bench, developed by Czech-Bayern cross-border laboratory of smart grid for testing different scenarios of P2P consumption-production analyses and security tests.

## ACKNOWLEDGEMENT

Smart grid in rural areas and SMEs – INTERREG V (no.144).

## REFERENCES

- [1] A. Berl et al., "Virtual Energy Information Network: A Resilience Perspective," *OVE e&i Elektrotechnik Informationstechnik*, vol. 130, no. 4-5, pp. 121–126, Springer, July 2013.
- [2] J. Murkin et al., *Enabling peer-to-peer electricity trading*.
- [3] R. Alvaro-Hermana et al., "Peer to Peer Energy Trading with Electric Vehicles," *IEEE Intelligent Transportation Systems Magazine*, vol. 8, no. 3, pp. 33-44, 2016.
- [4] J. Matamoros, D. Gregoratti and M. Dohler, "Microgrids energy trading in islanding mode," in *2012 IEEE Third International Conference on Smart Grid Communications (SmartGridComm)*, 2012.
- [5] C. Zhang et al., "Review of Existing Peer-to-Peer Energy Trading Projects," *Energy Procedia*, vol. 105, pp. 2563-2568, May 2017.
- [6] L. Marshall et al., "Allocation Rules and Meter Timing Issues in Local Energy or Peer to Peer Networks".
- [7] N. Capodieci et al., "Smart Meter Aware Domestic Energy Trading Agents," Jan. 2011.
- [8] P. Wegierek and M. Konarski, "The temperature effect on measurement accuracy of the smart electricity meter," Poland, 2016.
- [9] DigitKey, *Accurate Power Measurement in Smart Meters Part I*.
- [10] "OTE," 2018. [Online]. Available: <http://www.ote-cr.cz/>.
- [11] "Power Exchange Central Europe," [Online]. Available: <https://www.pxe.cz/>. [Accessed 2018].
- [12] "E.ON pricelist 2018," [Online]. Available: <https://www.eon.cz/-a119343---WPHOMJe-/cenik-komplet-elektrina-i-k-1-11-2017-distribucni-uzemi-eon-distribuce-2018-pdf>. [Accessed 2018].
- [13] D. J. Heidjann, "Electricity price composition (original title: Strompreis Zusammensetzung)," 2017. [Online]. Available: <https://www.stromauskunft.de/strompreise/strompreis-zusammensetzung/>. [Accessed 2018].
- [14] OpenUtility, "Piclo whitepaper," 2016. [Online].
- [15] RiberaSolutions, "IoT and Smart Grid," [Online]. Available: [www.riberasolutions.com/iot-and-smart-grid](http://www.riberasolutions.com/iot-and-smart-grid). [Accessed 2018].

# An Easy-to-use, Scalable and Robust Messaging Solution for Smart Grid Research

Ferdinand von Tüllenburg, Jia Lei Du, Georg Panholzer

Salzburg Research Forschungsgesellschaft mbH, Salzburg, AUSTRIA,  
email: {ferdinand.tuellenburg, jia.du, georg.panholzer}@salzburgresearch.at

**Abstract:** Smart Grids are characterized by tight coupling and intertwining between the electrical system and information and communication technology. Due to this, application layer messaging systems are regularly required for many Smart Grid applications. Especially in research messaging solutions are setup from scratch. In this paper we propose a generic and easy to setup message oriented middleware (MOM) solution providing robust and scalable messaging.

**Keywords:** Smart Grid, Messaging API, Middleware

## I. INTRODUCTION

Future electrical power systems will be characterized by a new control paradigm: Decentralized controllable power sources such as batteries, wind generators, and PV systems on production side and controllable loads on consumption side will be constantly monitored and operated depending on the current grid state in order to increase overall efficiency and ensure power quality. Part of this development is the augmentation of the electrical system with information and communication technology (ICT).

For the sake of data transfers between entities of the ICT subsystem usually a messaging solution is required providing an infrastructure to distribute messages correctly between instances also including definitions of message formats.

Especially for research projects in the field of Smart Grids, it is apparent that solutions for data transmission systems are redeveloped every time – a time consuming task when considering that the data transfer system is usually of minor priority. Due to this, an easy-to-deploy and (re-) usable messaging solution can be seen as a valuable contribution to Smart Grid research.

In this paper we introduce our messaging solution following the concept of a message-oriented middleware (MOM). As these features are regularly required in Smart Grid scenarios, the proposed solution provides robust, reliable, scalable and secure data transfers. All these without losing focus on ease of use and deployment, as well as applicability in various Smart Grid scenarios. In addition to this, we propose an application programming interface (API), which can be easily used and integrated in the communicating software components.

MOM solutions in general provide advantages for smart grid communication with respect to required communication capabilities (e. g. group communication) high scalability and high performance [1]. Additionally, one important beneficial aspect of using MOMs is that agents can focus on their key tasks of processing information, while the MOM handles

issues regarding security, performance, scalability, reliability and robustness of sending and receiving messages.

The paper shows the application of the messaging solution in context of an agent-based flexibility trading application.

## II. RELATED WORK

In context of messaging systems for Smart Grid application especially solutions based on XMPP are often used [2]. Although, XMPP is a flexible solution also following a MOM approach, it has weaknesses with respect to ease of deployment and configuration as well as implementation especially with respect to required aspects such as reliability. One example here is OpenADR[3]. Recently, with FIWARE, an open source platform is available which provides a large set of application programming interfaces (APIs) for a large variety of applications also providing a messaging solution for Smart Grids. However, the platform is extremely complex to setup and operated, thus lacking of ease-of-use.

In a more general sense, several MOM solutions are available. Those, however, have not been used widely in the field of Smart Grid applications. Particular types of MOMs use a (distributed) message broker as central hub for information interchange. Every sent message passes the messaging broker, which executes specific operations such as persisting, queuing or translating on each message. Especially with respect to reliability and robustness broker-based messaging has certain benefits such as life-time decoupling, state recovery, or guaranteed delivery [4].

Research and industry have developed several broker-based MOM systems providing similar services in general but have differences in their operational details and their specific focus. The solution ranges from research driven developments for certain use cases such as DoubleDecker [5] used for transmitting computer network status and monitoring data within a software-defined networking infrastructure, via solutions focusing mainly on high throughput such as ActiveMQ<sup>1</sup>, RabbitMQ<sup>2</sup>, and Apache Kafka [6] or lightweight and easy-to-use solutions such as NSQ<sup>3</sup> or NATS<sup>4</sup> towards totally cloud based solutions like Amazon SNS+SQS or the Microsoft Azure Platform. Despite of the fact that the usage of MOM has been encouraged also

---

<sup>1</sup> <http://activemq.apache.org/>

<sup>2</sup> <http://www.rabbitmq.com/>

<sup>3</sup> <http://nsq.io/>

<sup>4</sup> <https://nats.io/>

by others, apparently most Smart Grid solutions do not follow this approach.

### III. AGENT-BASED FLEXIBILITY TRADING

The context in which the proposed messaging solution is shown is a flexibility coordination scenario. Here, power equalization is achieved by orchestrating flexibilities of individual assets - particularly batteries, photovoltaics, and industrial loads. This means that assets adapt their production or consumption either by pre- or postponing or variation of amount. The flexibility coordination is implemented in a market-based approach, where flexibility offers are submitted towards a central market platform, where an optimal assignment of flexibilities is evaluated and corresponding control signals are sent towards flexibility suppliers.

The flexibility trading architecture is designed as hierarchical multi-agent system. Certain types of agents are situated at one of four logical layers fulfilling a specific task at this layer.

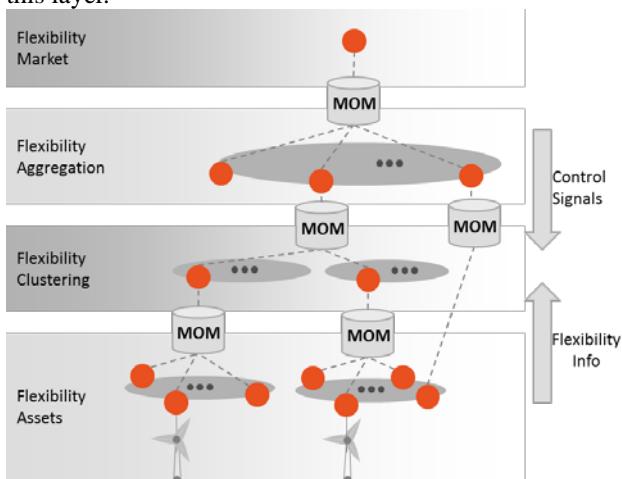


Fig. 1. Overview of the scalability-focused flexibility trading hierarchy including MOM instances.

Figure 1 shows the layered agent-architecture. At the bottom layer (Flexibility Asset Layer) reside device agents providing a common control interface of power grid hardware (e.g. a photovoltaic system or a battery) to the flexibility coordination system. Additionally, device agents implement a device-specific interface in order to read status information (such as state of charge of a battery) directly from the device or send control commands (e.g. charge or discharge) towards the device.

Within the hierarchy, each device agent may be directly connected to exactly one agent at the cluster layer or aggregation layer. Both are serving the purpose of condensing and aggregating flexibility potentials of assets. The difference between both layers is that the cluster agents group assets belonging logically together (e.g. situated in one facility) while aggregator agents group whole clusters or individual assets due to contractual (business) relation. Aggregator agents trade flexibilities at the market on behalf of asset owners.

At the market platform the aggregator bids are processed and flexibility potential is exchanged with respect to an optimal grid operation. A certain set of flexibilities is selected

at the market platform from the transmitted bids with respect to the current grid state.

From perspective of the information flow, in direction from device agents up to the market (upstream) flexibility information in form of device states or flexibility bids is transmitted. In the opposite direction (downstream), control signals for flexibility scheduling and activation are transmitted.

The deployment of MOM instances within the system architecture is done as shown in Figure 1. Instances are deployed at each border between two vertically neighbored layers for each combined group. Thus, the Kafka instances are provided and individually configured by either a cluster owner (e.g. a facility operator), an aggregator participating in flexibility trading and the market platform operator. This approach takes account of security, robustness, scalability and also control of system. The MOM providers at each level have better knowledge about connected agents, for instance by exactly knowing device types (e.g. for cluster owners) or contractual relations (at market or aggregator level). Additionally, the number of connected agents at each Kafka deployment is limited to a small amount compared to a centralized solution. Both aspects are especially important for large scale (e.g. pan-European) deployments where thousands of agents participate in flexibility trading.

### IV. KAFKA-BASED MESSAGE BROKERING

Due to the general similarity of available MOMs, the envisioned flexibility trading messaging solution would be potentially realizable with any of the MOMs named in the related work section. Thus, the selection of one the solution suitable for most messaging scenarios appearing in Smart Grid field is considerably difficult. In this project it was decided to implement our messaging solution on basis of Apache Kafka mainly due to some technical aspects, which seems making Kafka more suitable to the needs for Smart Grid messaging, compared to ActiveMQ and RabbitMQ solutions. The second reason for choosing Kafka is its proven usage in many large-scale productive environments.

Apache Kafka is designed as a distributed streaming platform following the principle of a broker-based message oriented middleware. Kafka's goal is to provide a means for high performance processing of continual sequence of input data (data packets) and its main design goals are reliable data transfer, fast processing, scalability, and fault tolerance [6]. These are important features for a Smart Grid messaging solution with respect to potentially many participating systems and overall criticality of such systems.

From a coarse-grained view, a Kafka-based distributed system consists of three principal components: The first component, the message producer, is creating streams of data that are read and/or processed by one or more message consumers (second component). The third main component is the Kafka messaging broker, which is mainly responsible for storing all messages until they are received by the message consumer. A Kafka message consists basically of a key/value pair for information encoding, a time stamp, and addressing information given by the core concept of message topics. A topic basically provides a name for a category of messages with a certain type. Using that name a message producer can

publish a message of that type. On the other end, consumers can subscribe to messages of that type also using the topic name. Each topic may have multiple consumers. The communication protocol used between Kafka clients and brokers is significantly more efficient than AMQP used by RabbitMQ and ActiveMQ. This gets in particular obvious when looking at a performance evaluation comparing latency and CPU load when batches of messages are processed<sup>5</sup>.

Several aspects specific to the design of Kafka are important to achieve the requirements specifically stated for the flexibility-trading application and many other Smart Grid applications. With respect to scalability of the system Kafka implements an efficiency-focused way of message handling by introducing the concept of partitions, being a particular difference to the queue implementations of RabbitMQ and ActiveMQ.

In order to provide high performance when messages are published or consumed, Kafka uses so-called partitions. Partitions can be seen as message sub-queues within a single topic, where published messages get stored in an immutable first-in-first-out order. Each partition is usually managed by one instance of a Kafka cluster potentially running at a dedicated host. In that way, write and read performance of a topic is substantially increased through allowing parallel writes and reads by multiple consumers and / or producers (compared to RabbitMQ, and ActiveMQ essentially only supporting concurrent reads). Published messages get sorted into exactly one certain partition of the topic - either directly specified by the producer or in a round-robin manner controlled by Kafka. For message consumption, basically, a consumer polls a topic and retrieves the messages of any of the partitions. Furthermore, in order to provide a means of load balancing and parallel processing at receiver side, consumers may be organized in consumer groups (in similar way supported by RabbitMQ and ActiveMQ), which lead to direct mapping between consumers within that group and certain partitions: Each consumer of the group retrieves the messages of one specific partition. Parallel reads and writes minimize the additional latency induced by message brokering in general. Using that approaches Kafka was able to achieve a throughput of 500 K Messages/sec [6].

With respect to fault-tolerance of the Kafka messaging system, each partition might have replications on other Kafka instances providing a means of redundancy. One of the replicas is considered as leader, where messages are stored first and afterwards replicated to other instances. Only after the message has been hard-drive-persisted at each replication, subscribed consumers can pull the message. This message persistence also provides a means for fault-tolerance in case of system failure of an Kafka instance. Together with a configurable retention time, specifying how long messages are stored in the Kafka system, this also provides a means of system recovery for message consumers, which then have the possibility to restore their internal state after a system failure by polling old messages from the respective topics. Similar functions are also supported by RabbitMQ and ActiveMQ, however, the particular page cache hard-drive persistence of

<sup>5</sup><https://dzone.com/articles/message-brokers-in-indirect-communication-paradigm>

Kafka outperforms other approaches resulting in lower latency and higher throughput [7].

In terms of security, Kafka provides TLS-secured connections between consumers/producers and the broker systems, and additionally access control lists (ACLs) specifying access rights (read, write) a consumer or producer has for each topic.

## V. CONFIGURATION AND USAGE OF THE MOM SOLUTION

Extending the proposed Kafka deployment, an easy-to-use messaging middleware client API for inter-agent data transfers was developed, aimed to be used by software component implementers. Currently, the API is available for Java and Python. Additionally, an efficient message serialization system based on Apache Avro [8] is included.

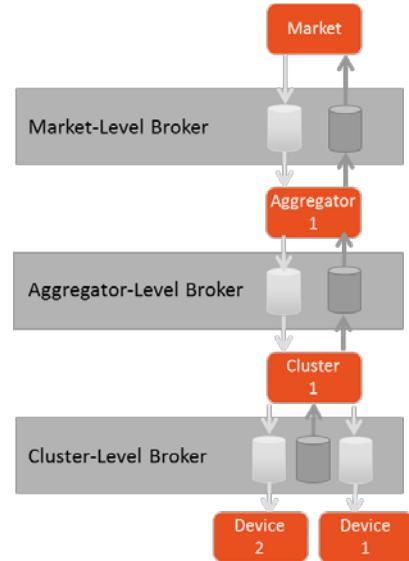


Fig. 2. Topic configuration for the flexibility trading use case.

The main design goals of the messaging client API were the provisioning of a high-level abstraction of Kafka messaging details such as creating and maintaining Kafka connections, topic and partition management, TLS key management as well as message encoding and serialization. This allows agent implementers to focus on developing agent logic despite of concentrating on messaging details. While the API and configuration is suitable for any arbitrary topology, we showcase the setup for the layered agent hierarchy as used for the flex trading use case. The corresponding topic configuration is shown in Figure 2.

Essentially, each agent uses one topic for incoming messages from lower layers (flexibility information), while top-down messages (control information) is delivered using private topics only readable by the addressed agents. This is chosen because of data security reasons but also in order to minimize the number of actively maintained connections for each agent (maximum of three in this case).

Before an agent is able to send or receive messages it must be registered at the Kafka broker including exchange of necessary security credentials. Within the agent hierarchy, each agent has connections to one or two Kafka brokers - one

broker for upstream data transfers and/or one for downstream. The messaging API provides the functionality to correctly connect an agent to the Kafka messaging broker. For this, however, a minimal set of information needs to be provided in advance for each connection an agent has to another agent in form of a simple configuration file which is given in Listing 1.

```
connection.name=aggreCon
agent.id=cluster-1
broker.address=192.168.100.135:9093
topic.in=cluster-1-in
security=TLS
```

Listing 1: Configuration file specifying settings for an agent connecting another agent.

Besides an unique agent identifier (`agent.id`) and a connection name (`connection.name`), the addressing information of the Kafka instance (`broker.address`), and the security configuration needs to be specified (`security`).

The API provides simple methods for sending (`send(Message)`) and or receiving messages (`receive()`) from other agents. The API user must not be aware of correct topic names and/or partitions or acknowledging received messages. The usage of the API is shown for cluster agent `cluster-1` in Listing 2 for the Java version:

```
//...
//instantiation of a connection object
Connection aggreCon = new Con("aggreCon");
//...
//create a MessageObject
Message request = new Message("Hello");
//reliably send message to the aggregator
aggregatorCon.send(msgObj);
//receive all messages from the aggregator
Message[] answers = aggreCon.receive();
```

Listing2: Instantiation and usage of a Connection object.

First, a `Connection` object is instantiated describing a connection to a particular broker instance as specified in the configuration file (which is automatically read when instantiated). Second, a message object is created. Third, the message is sent using the `send()` method provided by the API by specifying the addressee of the message. After that the user can be sure, that the message will be reliably delivered to the receiver. Calling the `receive()` functions fetches all messages stored in the corresponding topic (`topic.in`). Resolution of actual topic names as well as making sure that messages are sent and delivered is hidden by the API.

While configuration and usage of the messaging API seems to be simple, some remarks has to be made:

If (as in this case) private topics are used, the potential problem occurs that an upper-layer agent must know the names of any lower-layer agent prior to send messages. In case of our flexibility-trading application, however, this shortcoming is alleviated as an upper-layer agent is never required to initially contact a lower-layer agent. This situation is now exploited by inserting the ID of the sender agent into each application-level message. This allows an upper-layer

agent to derive the correct private topic name.

With respect to scalability and efficiency of data transfers, Apache Avro<sup>6</sup> [8] has been used for serializing and de-serializing application level messages into highly compact binary representations. The basic concept of Avro is using schema files defining how application data is to be serialized into binary data, and how binary data is to be interpreted by a reader. In order to achieve efficient data transfers, the schema files are not part of the actual binary data.

## VII. CONCLUSION AND FUTURE WORK

This work goes into details when MOM is applied in context of a flexibility-trading application as an example for applications in field of Smart Grids. These applications usually require scalable, fault-tolerant, and secure and reliable data transfers. We propose to use an Apache Kafka based messaging oriented middleware, embedded into a hierarchical system architecture consisting of various types of agents. It is shown, that several aspects of Kafka provide a beneficial basis in order to build a viable messaging solution. Additionally, we also provided details on a messaging API hiding much of the complexity of the messaging system and supporting implementers of agents.

As next step, this mainly conceptual work will be extended by applying and validating the messaging solution in a pan-European testbed. Here, the main focus will be laid on fault-tolerance and performance evaluations with respect to latency, throughput and also fault-tolerance. In this context we plan to investigate impacts of different Kafka deployment scenarios in comparison to the currently chosen multilayer deployment.

## REFERENCES

- [1] M. Albano, L. L. Ferreira, L. Miguel Pinho, and A. R. Alkhawaja. “Message-oriented middleware for smart grids,” *Computer Standards & Interfaces*, vol. 38, Feb. 2015), 133–143.
- [2] P. Saint-Andre, “RFC6120 - Extensible Messaging and Presence Protocol (XMPP): Core.” *Internet Engineering Task Force (IETF)*, Mar-2011.
- [3] M. A. Piette, G. Ghatikar, S. Kiliccote, E. Koch, D. Hennage, P. Palensky, and C. McParland, “Open automated demand response communications specification (version 1.0),” 2009.
- [4] F. von Tüllenburg, G. Panholzer, J. L. Du et al. (2017): “An Agent-based Flexibility Trading Architecture with Scalable and Robust Messaging,” *Proceedings of the 2017 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, Oct. 2017.
- [5] Wolfgang John, Catalin Meirosu, Bertrand Pechenot, Pontus Skoldstrom, Per Kreuger, and Rebecca Steinert. “Scalable Software Defined Monitoring for Service Provider DevOps,” *IEEE*, Oct. 2015, pp. 61–66.
- [6] Jay Kreps, Neha Narkhede, Jun Rao, and others, “Kafka: A distributed messaging system for log processing,” *Proceedings of the NetDB*, 2015 pp. 1–7.
- [7] Vineet John, Xia Liu. “A Survey of Distributed Message Broker Queues,” *arXiv preprint arXiv:1704.0411*, 2017

---

<sup>6</sup> <https://avro.apache.org/>

# Identity Management as seen by Orchitech

Ing. Martin Čížek, MBA; Mgr. Karel Maxa

Identity management, or just IdM / IAM, is a computer security discipline that enables the right individuals to access the right resources at the right times and for the right reasons. That's what Gartner says<sup>1</sup>.

You'd probably expect a polished product description, features and magic quadrants now. But it's not like that.



## We are software engineers

We first came across IdM twelve years ago. We were developing custom components for IdM solutions, including business logic bridges translating the IdM product to the customers' view of the world.

We were astonished by how much effort our colleagues were spending on tuning the core system setup, troubleshooting and operations across dev/test/prod environments. For instance, a forgotten difference among environments took a day of investigation. All that with expensive products by leading software giants.

We decided that "**this just must be simple**" and finally made a decision to do IdM projects on our own. We chose five cornerstones to build on:



**Community open source** – we share our efforts with international community. And we are the lead developer of the core component Wren:IDM, see the next page.



**Framework approach** – no bloated products the customer must fit to.



**SCM & Release management** – components, scripts and even configurations are under release management. Finding differences among test and production is then as easy as git diff.



**Test-driven development** – unit tests are part of our deliveries even for **configuration**. Changing a rule is no longer a QA nightmare.

*The idea is simple: imagine a brand new router. Does it work as expected? Depends on dozens of rules in its tables. Our unit test sends an "emulated packet" and checks what happens.*



**Software engineering best practices** brought to system integration and business processes make our competitive advantage.

*To name a few: zero config instead of golden images, code reviews and "know why it works" approach instead of "works somehow" faith. Semantic data model in the DBs instead of meaningless genericity interpreted in the middleware.*

## Simplicity and reliability pay off

Let us show off a little bit. The host of this conference, the University of South Bohemia, achieved an 80% drop in support requests after deploying our solution. And it's not only our minimalistic user interfaces; we are also addicted to code deletion. Usually when we receive a buggy legacy script, our revised version has also 80% fewer lines of code and works flawlessly.

## All you need but not more

We develop all common IdM features<sup>2</sup> as well as many specific ones and we deliver them as a completed puzzle tailored to the organisation. The specific things include integration with central workflow portals, multitenancy<sup>3</sup> or effortless deduplication. *By the way, did you know that an exchange student at a Czech university may appear with three different birth numbers?*

And since our architecture supports it, we are now preparing another piece of the puzzle, GDPR records<sup>4</sup>, consent management and personal data auditing.

<sup>1</sup> <https://blogs.gartner.com/it-glossary/identity-and-access-management-iam/>

<sup>2</sup> See <https://idm.systems>.

<sup>3</sup> One IdM deployment handles several organisations separately, e.g. faculties or subsidiaries.

<sup>4</sup> General Data Protection Regulation, Article 30.

### Ing. Martin Čížek, MBA



Graduated in Computer Science at ČVUT. Received an MBA Senior executive degree with distinction at B.I.B.S. Currently CEO of Orchitech.

### Mgr. Karel Maxa



Graduated with honors in Applied Informatics at JU. Now works as a solution architect for Orchitech's IdM solutions.



## Wren Security and Wren:IDM, open-source digital identity management

Wren Security is an open-source community security suite. It adopted community projects formerly maintained by ForgeRock™, that usually have roots in Sun Microsystems' products.

Orchitech is a lead developer of Wren:IDM, which is based on the former OpenIDM project. We are also involved in other projects like Wren:AM, the OpenAM successor, or Wren:ICF for integrating various systems. We offer a supported community product with all the benefits of our vendor status and international community at the same time.

We chose Wren:IDM as the core component for building our IdM solutions. It has a simple design and it's not a one-size-fits-all system, as our "Framework principle" suggests.

IdM is hot these days, as the GDPR pushes all to ensure compliance on all levels including organisational security.

**We owe the community more involvement and we need your help.**

Please see below how you can join us.

### BECOME A DEVELOPER

Java / Spring, JavaScript backend, JavaScript frontend, Groovy, OSGi, PHP / Symfony

Are some of these technologies familiar to you? Join us and contribute to great things.

You'd enjoy a nice team in České Budějovice (Budweis) or Prague, learn new skills and become a pro in new areas.

### BECOME A PAID VOLUNTEER

Penetration testing, security analysis, occasional development

Small contributions also count. Need a meaningful thesis topic? This can be a perfect match for you.

Some of "our" theses were exceptionally successful, with the dean's prize as the top achievement so far. What's more, a reasonable compensation can be expected.

### BECOME A SYSTEM ENGINEER

Linux / Docker, Gitlab CI/CD or TravisCI, application servers, web servers and databases, cloud platforms

Do you like to be on top of how systems are created and how they operate?

Join us in any of the mentioned schemes.

Visit <https://orchi.tech/careers/>



### Our passion: student identity and benefits

We are the main technological partner of the international ISIC organisation. We're creating a global platform that manages over 20 million cards in dozens of countries and hundreds of student benefits. The platform has integrations with around 100 systems worldwide. We're also developing a similar local platform for the ISIC's Czech representative GTS ALIVE, who leases it to 25 more countries. See more at <https://orchi.tech/isic/>.

### What Wren:IDM users say...

*I like what I see as far as thoroughness with the project. I noticed a few other fork attempts, but Wren is the only trustworthy one to me. It might be slow-going, but you guys are responsive and straightforward.*

*We'd also considered other products, but they are too difficult to learn, too tailored to a common corporate infrastructure, or have only high-level documentation. In particular, while other products only suggest possibilities, with Wren:IDM, I have the tools and the info I need to get it done.*

— **Guy Elsmore-Paddock**, solution architect at a leading vendor of enterprise software for the US grocery industry