

Vol. 3. Issue 2, 2017

ISSN: 2453-7314

Central European Researchers Journal



A WORD OF WELCOME FROM THE EDITORS

Dear Colleagues, Readers and Authors,

We would like to recall that the Central European Researchers Journal has been created under the CERES project (ref.no.: 544137-TEMPUS-1-2013-1-SK-TEMPUS-JPHES). The project is successful finished in this year. The closing event of the project was the workshop CERES: Modern Experience on Young Researcher's Organization. This workshop was held at University of Žilina on July 5-7 in frames of the International Conference on Information and Digital Technologies (IDT 2017), <https://idt.fri.uniza.sk/>.

The Conference IDT 2017 provides a forum for presentation and discussion of research contributions covering the multidisciplinary field of information and digital technologies. The main aim of the CERES workshop was to bring together young researches, teachers from academy as well as industry working in all areas of information technologies. CERES workshop will be good start of participants in their research career.

The papers of this number of the journal have been prepared after the conference and workshop as the post-publications of the best presentations. Authors of these papers are PhD students, young research and experts. The journal team follows the goals of the project and supports publications of both young researchers and experts. We hope to continue the traditions formed by the project CERES in future.

With best wishes

Prof. Vyacheslav Kharchenko

Prof. Elena Zaitseva

Central European Researchers Journal. Volume 3. Issue 2

Editor-in-chief: Kharchenko Vyacheslav, Zaitseva Elena

Editorial Board: Androulidakis Iosif, Belotserkovsky Alexei, Bezobrazov Sergei, Cariow Aleksandr, Cimrak Ivan, Dmytrychenko Mykola, Drahansky Martin, Drozd Alexander, Frenkel Ilia, Filatova Darya, Frigura-Iliasa Flaviu Mihai, Kachurka Pavel, Khakhomov Sergei, Kor Ah-Lian, Koshkin Gennady, Lapitskaya Natalia, Levashenko Vitaly, Liauchuk Viktor, Lukac Martin, Lukashevich Marina, Matiasko Karol, Melnychenko Oleksandr, Nedzved Alexander, Oliinyk Andrii, Pancerz Krzysztof, Ram Mangey, Slavinskaya Elena, Stankevich Sergey, Subbotin Sergey, Tatur Michail, Vojnar Tomas, Volochiy Bogdan, Yakovyna Vitaliy, Zhadanos Oleksandr, Zhivitskaya Helena.

Address of the editorial office: Central European Researchers Journal - editorial, Faculty of Management Science and Informatics, University of Zilina, Univerzita 8215/1, 01026, Zilina, Slovakia, editorial@ceres-journal.eu

Each paper was reviewed by reviewers.

Publisher: JMTM, s.r.o., Sad SNP 8, 010 01, Zilina, Slovakia, publisher@jmtm.sk

Published biannually

ISSN: 2453-7314

December 2017

REVIEWERS IN 2017

Adzinets Dmitry, Belarusian State University of Informatics and Radioelectronics, Belarus
Androulidakis Iosif, University of Ioannina, Greece
Belotserkovsky Alexei, The National Academy of Sciences of Belarus, Belarus
Cimrak Ivan, University of Zilina, Slovakia
Drahansky Martin, Brno University of Technologies, Czech Republic
Drozd Alexander, National Polytechnical University Odessa, Ukraine
Frenkel Ilia, Sami Shamoon College of Engineering Bialik, Israel
Kachurka Pavel, Brest State Technical University, Belarus
Kharchenko Vyacheslav, National Aerospace University - Kharkiv Aviation Institute, Ukraine
Kor Ah-Lian, Leeds Metropolitan University, United Kingdom
Kostolny Jozef, University of Zilina, Slovakia
Kvassay Miroslav, University of Zilina, Slovakia
Kvet Marek, University of Zilina, Slovakia
Kvet Michal, University of Zilina, Slovakia
Levashenko Vitaly, University of Zilina, Slovakia
Liauchuk Viktor, Gomel State University, Belarus
Lukac Martin, Nazarbayev University, Kazakhstan
Lukashevich Marina, Belarusian State University of Informatics and Radioelectronics, Belarus
Maceková Denisa, University of Zilina, Slovakia
Matiasco Karol, University of Zilina, Slovakia
Nedzved Alexander, Belarusian State University, Belarus
Oliinyk Andrii, Zaporozhye National Technical University, Ukraine
Slavinskaya Elena, National Transport University, Ukraine
Stankevich Sergey, Scientific Centre for Aerospace Research of the Earth, National Academy of Sciences of Ukraine, Ukraine
Subbotin Sergey, Zaporozhye National Technical University, Ukraine
Vaclavkova Monika, University of Zilina, Slovakia
Varga Michal, University of Zilina, Slovakia
Volochiy Bogdan, Lviv Polytechnic National University, Ukraine
Yakovyna Vitaliy, Lviv Polytechnic National University, Ukraine
Zaitseva Elena, University of Zilina, Slovakia
Zhadanos Oleksandr, National Metallurgical Academy of Ukraine, Ukraine
Zhivitskaya Helena, Belarusian State University of Informatics and Radioelectronics, Belarus

CONTENTS

<i>Alexander Nedzved, Alexei Belotserkovsky, Pavel Lukashevich</i> Software Solution for Direct Seismic Prospecting Task On Parallel Resources	1
<i>Uladzimir Parkhimenka, Mikhail Tatur, Olga Khandogina</i> Unsupervised ranking of clients: machine-learning approach to define a “good customer”	10
<i>Serhii Leoshchenko, Andrii Oliinyk, Sergey Subbotin</i> Analysis and software implementation of the methods for extracting semantic proximity between the lexical units	16
<i>Tyurin Sergey, Prokhorov Andrey, Vikhorev Ruslan</i> FPGA LUTs for a Logic Systems	25
<i>Anastasiia Plakhtii, Vadym Shkarupylo, Sergey Byelikov</i> The system for control and accounting of working time	39
<i>Tatiana Menshikh, Ilya Piatrenka, Vitaliy Khazan, Stanislav Derechennik</i> Comparative Statistical Analysis of Pseudorandom Binary Sequences by 53-bit Internal State Size Generator	45
<i>Uladzislau Beizerau</i> Fostering Competences Development in Belarussian Higher Education	51
<i>Igor Olenchenko, Andrii Oliinyk, Sergey Subbotin</i> Information technology for video broadcasting	58
<i>Anatoliy Gusev, Vladimir Kovalenko, Dmitriy Kostiuk</i> Virtualization-based architecture for the operating systems history demonstration	65
<i>Aliaksandr Zavadsky, Vasiliy Shut</i> Adaptive Algorithm of Intersection Management Based on Road Marking Adjustment Optimization	73
<i>Olga Nedzvedz, Sergey Ablameyko, Huafeng Chen</i> Application of the Integral Optical Flow for Identification of the Cell Population Motion in the Microscopic Images	80
<i>Valeriy Marunich, Oleksandr Melnichenko, Vitalii Kharuta</i> Mobility and Public Transport Service: From Deurbanization to Decentralization	89
<i>Peter Krcho, Simona Rusnakova, Oto Mihalco</i> Visualization of the ultrasonographical scans in critically ill neonates.	95
<i>CERES Press Release</i> Workshop on CERES: Modern Experience on Young Researchers Organization	105

Software Solution for Direct Seismic Prospecting Task On Parallel Resources

A.Nedzved, A.Belotserkovsky, P.Lukashevich, A.Dosin

Abstract— Conceptual scheme of hybrid technology for solving the direct seismic task using parallel resources for normal incidence for plane-parallel layer of rocks is presented here. Proposed software architecture connects thin client part with HPC cluster on the basis of SKIF supercomputer. Resulting synthetic profile of seismograms allows to verify a proposed geological model visually in real time.

Keywords— seismic prospecting task, parallel resources, Python, MPI, cluster technology.

I. INTRODUCTION

One of important IT-task in geophysics is to create an intelligent computer environment for the formalization of professional knowledge in the field [1,2]. The developing of a conceptual model and tools helps to realize the process of preparing a geological model, gathering and analyzing data presented in collected materials. To implement such tasks, it is advisable to use computer interactive tools for modeling and design processes in the "human-computer" interaction by supercomputer facilities.

An improvement of known modeling methods is very crucial for the formalization of professional knowledge. Efficiency of technological processes and systems is supported by complexity and multivariate construction since organizing of computational processes is associated with the data processing when complex technological systems are designing. It requires interactive and graphical tools to organize the process. The choice of solutions is usually based on technological characteristics and parameters of the designed facilities [3-5].

It is known that such tasks are of high information complexity due to processed objects. It leads to the decomposition of the structures into component parts, with the separate processing of each part and their subsequent synthesis. High computational intensity and low level of formalization of initial data can be compensated with complexing of interactive tools with parallel computing.

II. DIRECT TASK OF SEISMIC EXPLORATION

The solution of direct task of seismic research on parallel resources for normal incidence for plane-parallel rocks layer is a good example for demonstration of application parallel resources of the cluster technology (such as SKIF supercomputer [7]). Direct task of seismic prospecting is the calculation of arrival times t and amplitudes A for a separate wave in a known seismogeological section. This task is solved when following data is known: density, depth of occurrence, size of certain geological objects (layers), speed of distribution of elastic waves and location, shape of a rock source [8,9]. A simple solution of direct dynamic seismic problems of inhomogeneous data is realized by solving of the following wave equation:

A. Nedzved, Belorussian State University, Minsk, Belarus, NedzvedA@tut.by.

A. Belotserkovsky, United Institute of Informatics Problems, Minsk, Belarus, Alex.belot@gmail.com.

P. Lukashevich, United Institute of Informatics Problems, Minsk, Belarus, pavel.lukashevich@newman.bas-net.by.

A. Dosin, United Institute of Informatics Problems, Minsk, Belarus.

$$\frac{1}{V^2} \cdot \frac{\partial A}{\partial t} = \frac{\partial^2 A}{\partial x^2} + \frac{\partial^2 A}{\partial y^2} + \frac{\partial^2 A}{\partial z^2},$$

where V – is velocity of wave, $A(t, x, y, z)$ – is magnitude of signal, propagation through a rock with coordinates (x, y, z) for different value of time t after initial stage.

The solution of the equation uses the boundary conditions. It can be implemented for simple rock models. In this way the kinematic problem is much simpler. It allows to determine the arrival time of a wave (direct, reflected, refracted, etc.) for a defined model. Such solution exists only for preset position of the source and the time of initiation of the elastic wave. Traditionally, the simplest result of solving of a direct task is obtaining by the hodograph equation. It is analytical expression for the function $t(x)$. As result the hodograph is a graph of time dependence of the arrival of a wave from the distance (x) . It is ranked from the source point to the point of reception.

Seismic research uses the method of reflected waves in a situation when the depth of rock layer is much bigger than the maximum distances between the source and the receiver in the case of normal incidence of waves. If the angles of incidence are not exceeding the critical value, the pulses of the exchange waves have the same shape as the pulses of the incident waves. Mixed and exchanged waves at normal incidence angle are not created. When decision is making, we generate a number of simplifying assumptions. However, the solution of the direct seismic survey task for normal incidence of waves in a thin-layered rock model improve studying of the wave characteristics of fields corresponding in real environments [9].

III. DATA PREPARATION FOR TRACE SIMULATION

The seismic geological model is constructed from seismic layers (seams). Its basic feature is the limitation in space by simple boundaries. It is interpreted as soles and roof of the layer. Such model differs from the enclosing layers with elastic properties. Physical properties of a homogeneous layer are constant. For gradient layers they have monotonic changings. In practice, homogeneous models of layers have usually priority in investigation. Thus, the simplest models of horizontal layers are mostly used. The sections of the sedimentary cover in many areas of seismic exploration are well approximated for them. For layered gradient rocks, practical applications are solved by models with gradients of vertical velocity that reflect the most general pattern of changes of elastic properties in sedimentary deposits [9].

Characteristics of the vertical coordinate of the layer determine the basic parts for the hodograph equation. These characteristics are determined from the lines of geological horizon. It limits the layer. Properties of lines are defined interactively by user at the “tablet” (main interactive windows for seismic data visualization) of the analysis of seismic profile, as shown in Fig. 1.

The line of horizon has following characteristics: name, velocity of wave propagation, density and absorption coefficient. The name of the horizon is an identification parameter, which is settled by the user when the model is prepared by researcher.

The density of the rock is also determined with additional geological prospecting data by the user. It includes characteristics that are obtained from analysis of a geological well. In this case the density is defined as the mass in a unit of volume for the solid phase of the rock. So, it affects to the velocity of wave propagation in the rock.

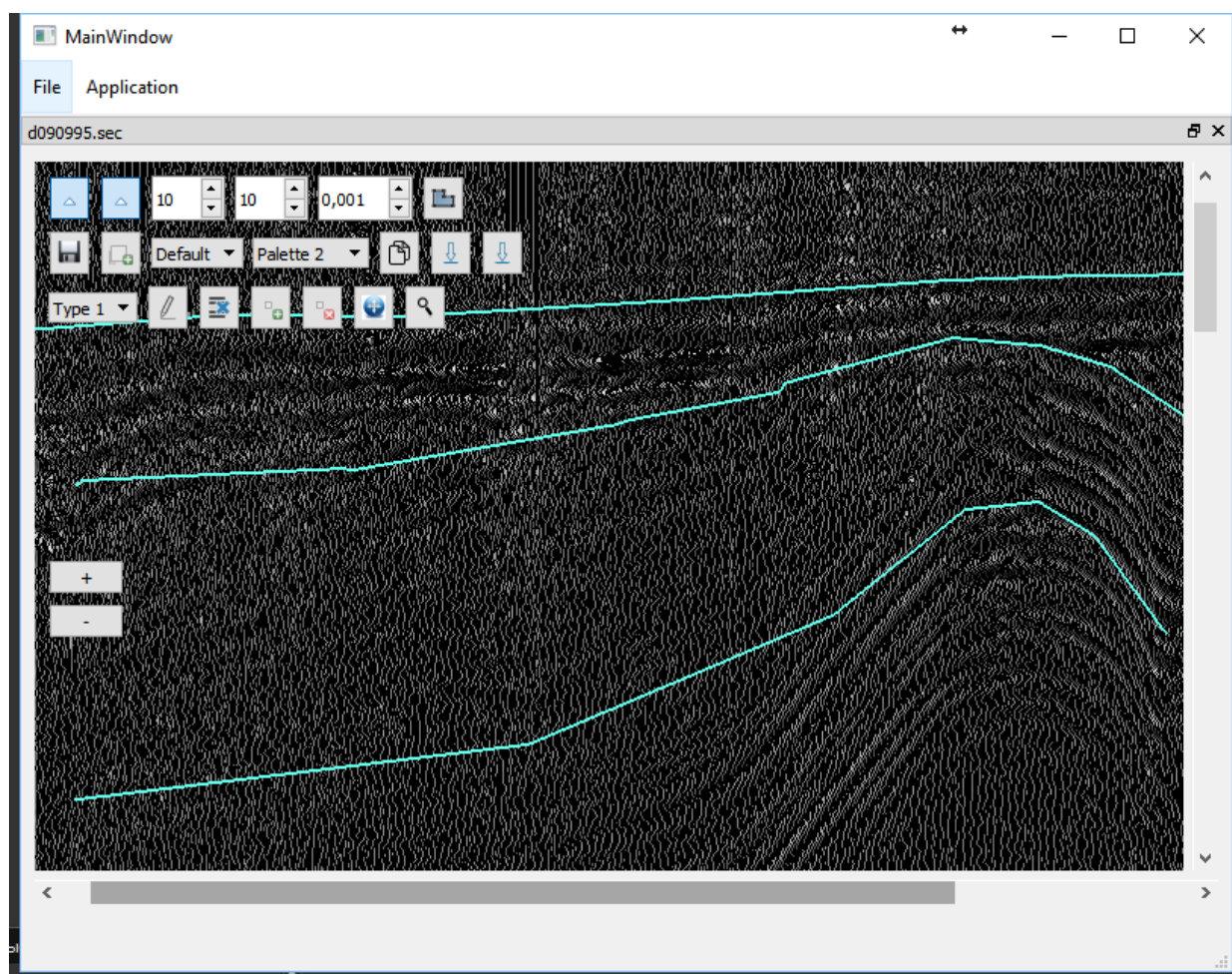


Fig. 1 Screenshot of the “tablet” with seismic profile where lines of limitation of layer is prepared

The time of the wave propagation in the rock is determined from point of its reflection in boundaries of the layer to the point of a recording sensor. In the time section, the reflecting horizon for bottom layer of the rock has a tracing time. Then, the depth from the horizontal boundary is calculated as the multiplication of half the tracing time to the wave speed, that is defined as average value of velocity of tracing from the boundary. The tracing time characterizes the depth of occurrence of the rock.

The absorption of seismic waves depends on the lithological composition and physical state of the rock. The absorption coefficient characterizes one of the physical properties of the rock and can be used for geological interpretation of seismic observations. In the algorithm for normal incidence for parallel plane layers, this coefficient is not used.

The next set of raw data is used to implement the solution of the direct seismic task at normal wave incidence. These are characteristics of impulse that are used to construct of syntetic trace. As a rule, it must correspond to the signal, which is used in geological research of the seismogram registration process. These impulses have different localization in the frequency-time domain. Therefore, the impulse is preliminarily chosen for calculation. It should look like as a research signal.

The simplest way to choose an impulse is a determination of the spectrum amplitude, which most completely coincides to the spectrum amplitude of the autocorrelation function of a real seismic trace for a zero-phase realization.

In our software, the following types of impulses are implemented for using as the initial seismic signal (impulse):

- Riker's momentum;

- Puzyrev's pulse;
- Sinc in Window;
- Decaying sinusoid;
- Damped cosine;
- Wavelets:
 - a) DOG;
 - b) WAVE;
 - c) MHAT ("Mexican hat");
 - d) Mople real (the real part of the complex Moplet).

In addition, the impulse can be simulated by calculation program. In this case, it is determined by a numeric vector or one-dimensional matrix. Basic impulses used in the system are shown at Fig. 2.

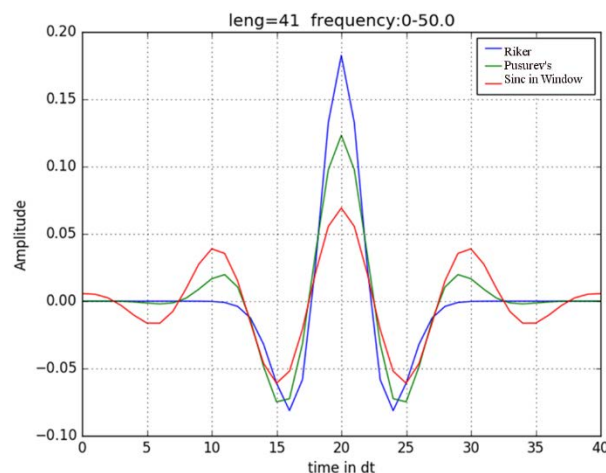


Fig. 2 Basic impulses for trace simulation

The next set of data determines the way of data processing for model trace simulation. It consists of a sequence of sets. These sets are describing the intersection points of the seismic trace and the lines of the horizons. Such set has digital data of characteristics and name of layer:

- the final processing time;
- step of discretization;
- impulse characteristics that determine its reading from a file or calculation set;
- wave characteristics for single and multiple reflection;
- the polarity of the mapping and the construction of the wave (right or left).

If it is possible to read a numerical series from the impulse section for its assignment, an impulse is formed from it. Otherwise, it is calculated on the basis of the characteristics of the seismogram. If the variable is equal to "All", all possible processing options are evaluated.

The last set of source data is the set of values of seismogram. This set includes general characteristics of traces and consists of the following data:

- The name of the seismic profile;
- The name of the picket or trace;
- The initial time for the trace;
- The final time for the trace;
- Step of discretization of the trace..
- Profile and picket names are string values.
- Other variables are characterized by numerical values. They are specified in the specific header before a set of discrete values, which correspond to the values of the trace.

In the software the special function is realized. It provides analysis of these data and constructs specific structure that allows us to divide data for parallel processing. Also it is shows in specific window with structures (Fig. 3).

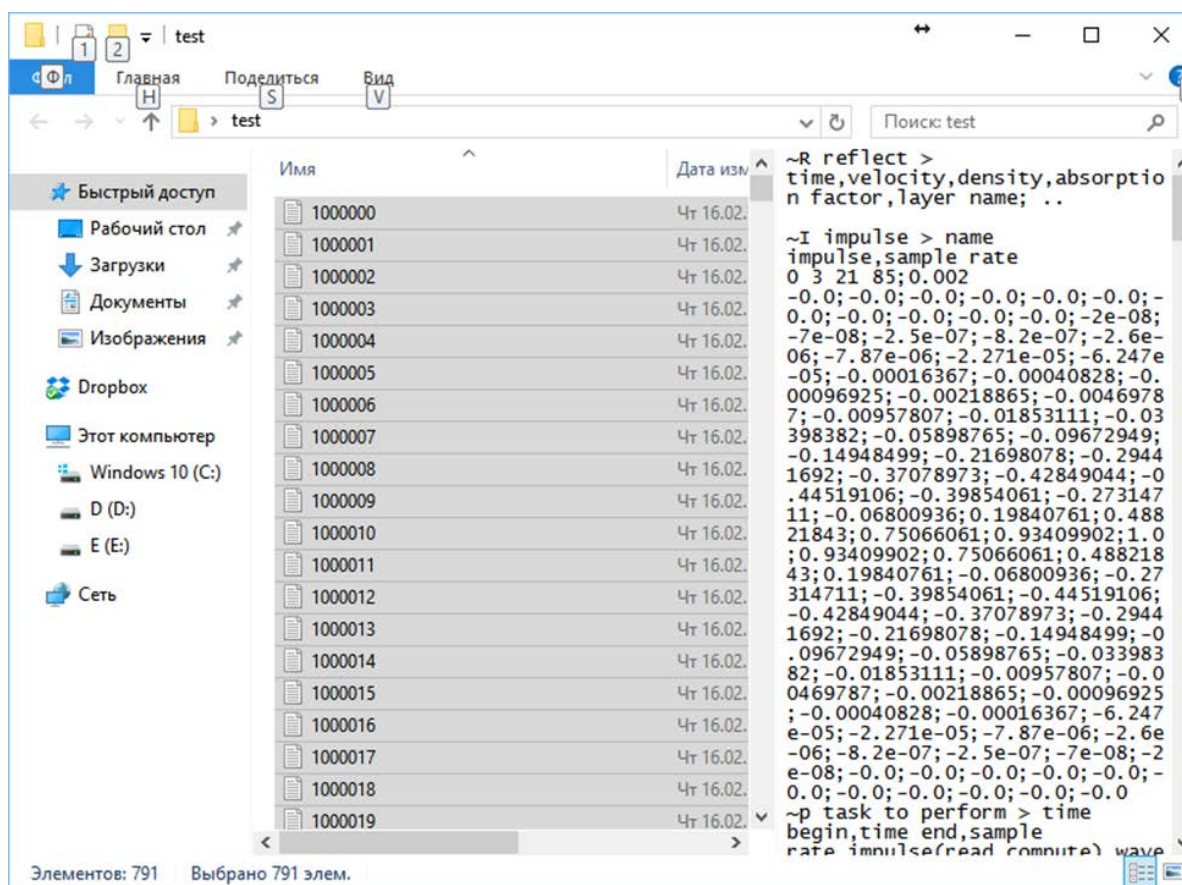


Fig. 3 Special window with data structure for parallel processing

This information describes only geological content that is used for construction of new profile of synthetic seismograms. It uses a special representation of the information, which is broken down into separate traces. As a result, the synthesis can be performed for each individual trace. Then the result can be combined in the profile of synthetic seismograms.

IV. PARALLEL SOLUTION FOR THE SEISMIC SURVEY TASK

Running a software for seismic prospecting on server will bring a delay of graphics activation due to a transfer rate of graphic content for seismogram. Using a network with common topology and average characteristics the visualization has meet a speed equal to 3fps. On the other hand, there are long delays in simulating of seismic traces, if the software is started from client side. Thus, the calculations of the traces simulation and interactive processing at graphics of seismograms profile representation can be executed on server and on client parts separately.

The basic problem of such realization is separation of information flows. The solution of the seismic task on parallel resources is executed at the computing architecture of SKIF family (cluster technology). It uses an operating system that based on the CentOS linux kernel. In such system the access to the cluster is organized via SSH protocol. To organize such access, the software used libssh 0.3 library.

The libssh library is designed to integrate supporting of SSH protocols into C source code. The library allows to run programs on a remote server, copy files through a secure connection. The library supports both standard scp and sftp transfer mechanism. It is implemented without external libraries, such as libcrypto (from OpenSSL) or libgcrypt. Therefore, such developing of software is easier.

Libssh implements server and client functions, supports SSH1 and SSH2 protocols, AES-128, AES-192, AES-256, Blowfish, 3DES encryption modes in CBC, the ability to work with RSA and DSS public key servers. It supports data compression, co-working with ssh-agent and the organizing several sessions within a single connection.

The QSshSocket class is also implemented to interact with the software “tablets”. It is designed to create an asynchronous software interface for data exchange (socket) using the SSH and SCP protocols. The class is specially implemented to use it with the Qt library, which is cross-platform and corresponds to logics of “tablets”. The class effectively uses the functions of the libSSH library.

The QSshSocket class is a child of the standard QThread class from the QT library. In addition to the standard features, it includes an additional set of constants and functions.

These functions allow the client and SKIF clusters [10] to interact through data flow (Fig. 4).

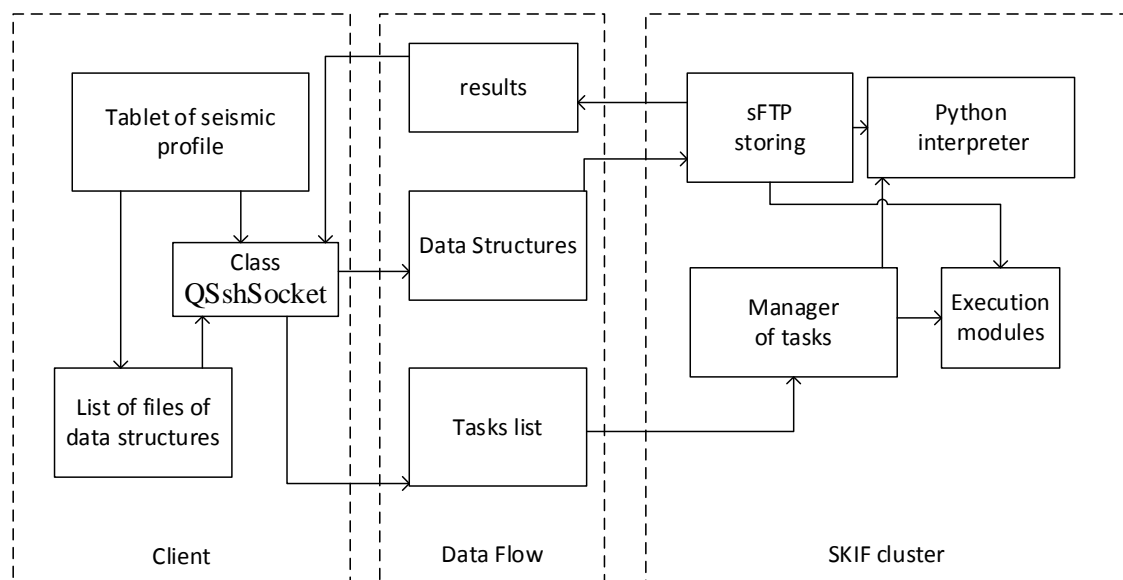


Fig. 4 Scheme of implementation of software for geological interpreting by simulation of direct seismic task

According to the scheme we divided the graphic and computational part. It allowed us to accelerate the interactive work with a graphical representation of the seismic profile.

A process management was realized by the Portable Batch System (PBS). The PSB package is designed to manage computers and system resources. It allows to increase an effectiveness of managing of calculation resources by refusing equipment downtime. PSB generates sets of task for execution, it stores, starts and controls of execution every task and sends information about implementation of calculation processes to user.

The clusters SKIF use free version of PBS – Torque, which is based on famous system OpenPBS. It used Maui package as free manager of tasks. PSB’s directives are transferring to computational system by separated command QSUB or through execution script. The software uses following features:

1. Starting of execution software from command prompt with argument, for example:


```
qsub -l walltime=00:20:00,nodes=3:mpp /opt/scali/bin/mpirun -machinefile nodes.sca \
$PBS_O_WORKDIR/python theogram.py 3806.txt.
```

This sample launches a python script to calculate a synthetic seismogram for separate picket.

2. Creating script of PBS for starting package of software and commands have parameters of tasks. They can be set by a string: “qsub <script_name>”, where <script_name> is name of script. For execution that string is generated by client part of software. This script is created by client part of software (“tablet” of seismogram profile). Commands from script are queued up as jobs in “dteam” (“batch” on the cluster).

In the simplest case, when a uniprocessor program is started on a Linux cluster, a script file is created with a set of commands as follows:

```
### PBS script to run one CPU task ###
```

```
#PBS -l walltime=00:10:00
```

```
#PBS -l nodes=1
```

```
#PBS -N python theogram.py
```

```
hostname
```

```
cd ~/test_pbs/
```

```
date >> ~/test_pbs/log
```

```
hostname >> ~/test_pbs/log-
```

```
./python theogram.py 3806-0001.txt
```

```
./python theogram.py 3806-0002.txt
```

```
./python theogram.py 3806-0003.txt
```

```
.....
```

```
.....
```

```
.....
```

```
./python theogram.py 3806-0748.txt.
```

This script initiates an environment of parallel processing by creating tasks’ sequences to calculate a set of synthetic seismograms. A few variants of calculation are to be started by the next command:

```
qsub -l walltime=00:20:00,nodes=3:mpp /opt/scali/bin/mpirun -machinefile nodes.sca \
$PBS_O_WORKDIR / python mergetheogram.py 3806-0001.txt 3806-0748.txt.
```

As a result, a new profile of synthetic seismogram is created. A seismogram for every picket is calculated as a separate process at separate node. The same sequence of command is used to calculate each seismogram with corresponding initial data for each picket.

Seismograms from pickets are processed on the server and the software generates a folder with solution which is necessary to verify. It allows to sort resulting data by picket number. Calculated data sequence is merging to general array of seismogram and representing as new synthetic profile. Several synthetic profiles for different phases or impulses can be built. These profiles are sent to client part of software by SSH protocol and shown in separate graphics windows (Fig.5).

V. MERGING SYNTHETIC SEISMOGRAM OF PICKETS TO PROFILE

The United Institute of Informatics Problems of the National Academy of Science has organized the “State supercomputer multi-access center” (SSMAC), which includes the computational capabilities of cluster supercomputer family. The high speed approach to the resources connected to telecommunication network of the National academy of Sciences of Belarus BAS-NET from scientific networks that is connected to BAS-NET with fiber-optic. The network interaction of remote users with SSMAC is organized by protocol SSL.

We have tested our software by three cases of equipment:

- Desktop workstation;
- SKIF cluster by Xming that is an X Server running on Microsoft Windows;
- Hybrid scheme where interactive graphics is realized on PC-client and calculation on SKIF supercomputer.

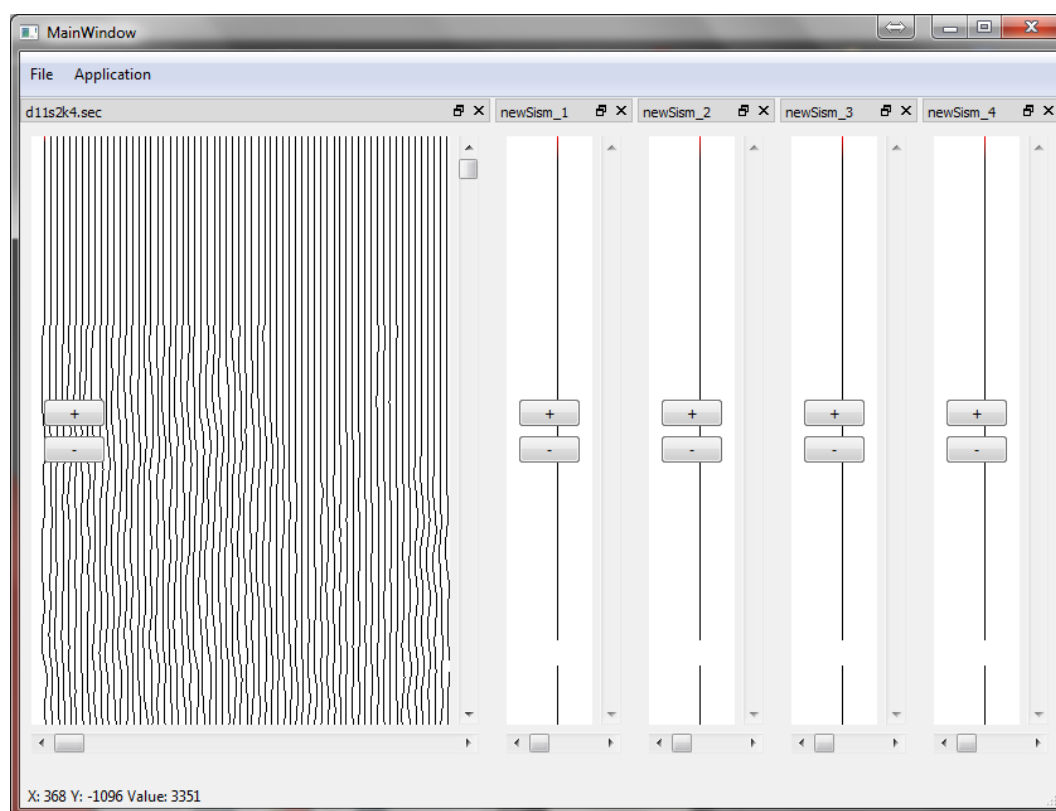


Fig. 5 The set of window with real and synthetic seismogram after seismic simulation.

The test was executed by specific sample. The source data for testing include real profile of seismogram, nine lines of geological model and sets of basic impulses. The profile of seismogram includes traces for 1240 pickets with 3000 discrete values. Geological model has about nine points of rocks layer for every picket. Sets of impulse include Riker's momentum. We use simple algorithm to solve a task for normal incidence for plane-parallel rocks layer. It is based on signal deconvolution.

We estimate two parameters: frame rate and speed of calculation. The frame rate is defined as count of frames per second (fps). This parameter allows to estimate properties of “tablet” of seismic profile.

The speed of calculation is defined through a time of script execution to solve a task for normal incidence for plane-parallel rocks layer.

The result of testing is represented in Table 1.

TABLE I
Units for Magnetic Properties

Type of platform	Execution marks		
	<i>FPS</i>	<i>Time of estimation (s)</i>	<i>Common mark</i>
Desktop	30	220	0.13
SKIF cluster	5	35	0,14

Type of platform	Execution marks		
	<i>FPS</i>	<i>Time of estimation (s)</i>	<i>Common mark</i>
Hybrid scheme	30	61	0,41

We define common mark of effectivity of software execution as relation FPS to time execution. Test show that hybrid technology is very effective for comfort of geological research of seismic data.

VI. CONCLUSION

In this paper, the hybrid technology is described for solving of direct seismic task at parallel resources for normal incidence for plane-parallel layer of rocks. The proposed software environment connects a “tablet” of seismogram profile for interactive modeling with calculations at server. It uses the SSH protocol for communication issue. The computational tasks are solved using the Python interpreter, which is located on the cluster of the SKIF supercomputer. The solution of the seismic problem is divided into tasks for each trace or picket for parallel execution at nodes of the supercomputer. For each task, a file and a command line are generated, which are transmitted by the SSH protocol to the cluster for execution. For finishing, the general solution of the task merged all result. As a result, a new synthetic profile of seismograms is created for visual verification of the geological mod

ACKNOWLEDGMENT

The research is supported by Joint Russian-Belarusian Scientific and Technical Programme “Research and Development of High-Performance Information and Computing Technologies for Increasing and Effective Usage of Resource-Intensive Hydrocarbon Resources of Union State (SKIF-NEDRA)”.

REFERENCES

- [1] Ponsa X., Masó J. A comprehensive open package format for preservation and distribution of geospatial data and metadata // Computers & Geosciences, vol. 97, 2016, pp. 89–97.
- [2] Vichalai C., Huy Giao P., Puttiwongrak A. Integrated analysis of high resolution seismic and geotechnical data using opensource software for an offshore drilling rig site // Symp. on Imaging and Interpretation, 2009, pp. 5-9.
- [3] Sato H, Fehler MC, Maeda T. Seismic wave propagation and scattering in the heterogeneous earth, 2nd edn. Berlin, Germany: Springer. 2012, P. 184
- [4] Virieux J, Operto S. An overview of full-waveform inversion in exploration geophysics. // Geophysics, vol. 74, pp. 2009, pp.1-26.
- [5] Graves, R. Simulating seismic wave propagation in 3D elastic media using staggered-grid finite differences. // Bulletin of the Seismological Society of America, vol. 86, 1996, pp.1091-1106.
- [6] Ablameyko S., Abramov S., Anishchanka V., Paramonov N., Tchij O. 2008. SKIF Supercomputers and Their Applications // Problems of Cybernetics and Informatics: proceedings of the Second International Conference. September 10 -12, 2008, Baku, Azerbaijan. – Baku, Azerbaijan National Academy of Sciences, 2008, pp.3-7.
- [7] Avseth, P, T Mukerji and G Mavko. Quantitative seismic interpretation. // Cambridge University Press, Cambridge, 2005, p. 183
- [8] Louie J.N.; Pullammanappallil S. K. Advanced seismic imaging for geothermal development // New Zealand Geothermal Workshop 2013 Proceedings. 2013. pp.1-7.
- [9] Laine J.M. Seismic interpretation and stratigraphy using opensource software // Proc.of Geological Society of America Abstracts with Programs: Vancouver Convention Centre-West, vol. 46, no. 6, 2014, pp. 356.
- [10] Nedzved A., Belotserkovsky A., P. Lukashevich Building the regional system for seismic data processing // Proc.on Information and Digital Technologies 2015 International Conference. Zilina, 2015, pp. 261-265

Unsupervised ranking of clients: machine-learning approach to define a “good customer”

Uladzimir Parkhimenka, Mikhail Tatur, Olga Khandogina

Abstract—Ranking of clients is a natural problem for every business. Though usually it can be solved by common sense and intuition of managers, in the case of a big business entity (e.g. global online stores), the problem becomes more complicated with obvious obstacles in derivation of fast and accurate solution. This article deals with the clients ranking problem using machine learning methodology.

Keywords—loyalty ladder; ranking; ecommerce; automatic marketing decision-making; machine learning; data mining & knowledge discovery; latent variable analysis.

I. INTRODUCTION

Every company, even the biggest one, faces the problem of limited resources, which is the central economic problem in principle. For example, in a product strategy a company has to decide what product portfolio composition should be optimal, what products or models to produce (and in what quantities), what products or models to discontinue, while there is obvious limitation of production capacity. In an investment strategy, a company deals with defining priorities of assets to invest in, given a limited budget of financial resources.

The same problem a company has in relation to a pool of clients. There is a need to know what clients are worth of marketing activities and efforts and what aren't. Such knowledge could be used for tailoring individual loyalty program for each client depending on his/her 'worth' and 'quality'; introduction of personal managers for 'good' clients; cutting down marketing efforts and costs on 'bad' clients; introduction a queue policy in provision of services, etc.

Furthermore, a company could not need only a qualitative solution here ('bad' or 'good' client), but a quantitative one. A company needs to prioritize its clients, thus making a ranking.

Ranking is sorting of objects in accordance to one or several criteria. The result of ranking is a hierarchy: from the 'best' (biggest, most valuable, tallest, etc.) to the 'worst' (smallest, least valuable, lowest, etc.). Such an approach is widely used in the society, e.g. a music chart or university rankings.

Ranking is about comparison of several objects with one another; it reveals preferences of a decision-maker and is closely related to choice modelling.

In the case of objects with many characteristics, ranking is a way of simplifying complex relationships between objects. It could result in losing of some information about these objects, but gives a very clear and simple 'picture' as a set-off against this lost.

In the business literature, there are many approaches to the problem of clients ranking, usually in the form of assigning clients to several (not too many) categories based on the history of their interactions with a company. For example, companies distinguish a few 'platinum', 'gold', VIP-, key account-, top-clients contrary to a majority of 'other' or 'normal' clients.

The key question here is what exactly constitutes the notion of a 'good' and 'bad' client for a company. There is no definite answer, though financial (e.g. sales volume or profit), indicators seem to be main contributors here, while strategic goals (e.g. competitive and positioning strategy) are also used, but play less significant role.

This article looks at the perspectives of using machine-learning methodology to solve the problem of clients ranking in unsupervised manner: without any expert opinion and learning sample.

II. TRADITIONAL APPROACHES TO CLIENTS RANKING

In the business literature, there is no any universally recognized method for ranking of clients. Though the necessity of distinguishing ‘better’ client over ‘worser’ ones is widely taken for granted, e.g. in the framework of the well-known key account management (KAM) methodology or within the concept of loyalty ladder, there is a variety of different approaches to the problem.

Pareto Law states that 20% of clients generate 80% of all sales volume (or profit) [1]. Therefore, the natural consequence of such a statement is to divide clients in two groups (in proportion approximately 20:80) and focus main efforts on the smaller, but more ‘worth’ part.

ABC analysis, another very popular in the business literature approach to the problem, could be considered as a natural follow-up of the Pareto Law. Its primary focus is inventory [2], but clients can be and are analyzed as well [3, p. 15]. The main idea of this approach is division of all objects into three groups in accordance to their share of sales volume or profit or other outcome (e.g. value added). The *A* group consists of objects that generate approximately 80% of the outcome. To the *B* group belong objects that generate next 15% of the outcome. The *C* group comprises objects that generate the last 5% of the outcome. Theoretically expected that the size of *A* group would be approximately 20%, *B* – 30%, *C* – 50% of the whole clients population.

There are several extensions of classical variant of ABC analysis, for example multi-criteria analysis [4].

XYZ analysis is considered to be another one, dynamic extension of the static ABC analysis [5]. This approach is oriented in defining what objects (usually inventory, but in this case – clients) show stable ‘behavior’ over the period of time and what don’t. The ‘stability’ is measured by statistical variance.

RFM analysis is a popular and traditionally incorporated in CRM systems approach [6, 7, 8]. It aims not directly on ranking, but on clustering of existing clients in accordance to three dimensions: Recency – time from the last purchase, Frequency – number of past purchases during certain period (e.g. year or quarter), and Monetary – volume of purchases during the same period. Nevertheless, it gives some insights what clients are ‘better’ than others, at least within a specific cluster.

Customer Profitability Analysis (CPA) [9] focuses on measuring profitability of a customer as a main indicator for policymaking. Its logical successor, Customer Lifetime Value approach (CLV) [10, 11], aims at calculating the net present value of a single client (customer) given expected average purchase volume, its frequency, duration of customer “lifetime”, cost of capital (discount rate). Ranking is a natural consequence of this analysis (CPA or CLV), though it is never meant explicitly.

III. ‘GOODNESS’ OF A CLIENT AS A SYNTHETIC LATENT VARIABLE

If we look to the essence of the above-mentioned approaches and switch from the ‘business language’ to a formal one, we could generalize all approaches as a simple weight-and-sum ranking system:

$$Score_i = \sum_{j=1}^n w_j \cdot x_{ij} \quad (1)$$

where $Score_i$ – an estimate of the ‘worth’ (‘quality’) of *i*-th client that is used further for sorting from maximum to minimum to get a final ranking;

w_j – a ‘weight’ of the *j*-th characteristic of a client;

x_{ij} – a value of the *j*-th characteristic of a *i*-th client;

n – number of characteristic of a client taken into account for ranking.

For simplicity, we skip other two approaches to specify a ranking model (pairwise and listwise).

In traditional methods (see section II of this article) viewed from the perspective of (1), in order to get final ranking one has to specify: (i) set of n characteristics of a client (observable, measurable variables) as partial indicators of a client's 'worth' and 'quality'; (ii) values of w_j as 'weights' and 'importance' of a specific characteristic.

In general, characteristics of clients can be of a great variety, e.g. in the case of online stores visitors:

- purchase volume in monetary terms during the specified period;
- purchase volume in-kind during the specified period;
- profit generated by a client during the specified period;
- number of previous visits during the specified period;
- expected Customer Lifetime value;
- RFM-metrics;
- behavioral traits (e.g. number of pages viewed at online store or number of goods put in a 'basket');
- engagement metrics (e.g. number of 'likes' or 'reposts' of relevant content in social media);
- etc.

The chosen set of characteristics and their 'weights' reflects experts' understanding of situation and preferences of owners and top managers of a company through.

Variable $Score_i$ means the 'worth' ('quality') of the i -th client. This is unobservable variable, that can not be measured directly.

In statistical terms, it is rational to use latent variable approach [12] or latent synthetic categories approach [13]. We treat both approaches as equivalents. Both of them are oriented toward constructing of a directly unobservable variable as a function of directly observable variables. Formula (1) fits this definition in full.

Usage of latent variable approach allows to make rankings without knowing preferences of experts and without any learning sample.

IV. PROCEDURE OF UNSUPERVISED RANKING OF CLIENTS

Latent variable ranking of clients should include several steps.

Step 1. Defining a set of variables (characteristics of clients). This can reflect specific preferences of decision-makers (e.g. profitability over sales volume) or can be just a set of all possible variables tracked by an information system of a store relevant directly or indirectly to 'worth' or 'quality' of a client. Some examples are given in the Section III of the article.

Step 2. Getting data. It means getting $n \times m$ values x_{ij} for all m clients with n characteristic.

Step 3. Scaling variables to the interval [0, 1]. This transformation can be done by classical approach:

$$Z_{ij} = \frac{x_{ij} - \min_j x_{ij}}{\max_j x_{ij} - \min_j x_{ij}} \quad (2)$$

In the case of reverse relationship between a variable and 'worth' and 'quality' of a client (e.g. number of returns of goods), should be used another formula:

$$Z_{ij} = \frac{\max_j x_{ij} - x_{ij}}{\max_j x_{ij} - \min_j x_{ij}} \quad (3)$$

Such an approach does not take into account occurred statistical distribution and possible data outliers, but a data analytic should be aware of that and make corrections if needed.

Step 4. Making a principal component analysis. This would give a set of principal component scores (client's 'coordinates' in principal components space) – s_{il} , where i denotes the i -th client, and l denotes l -th principal component.

Step 5. Defining the number of principal components for ranking function. Following the approach of Aivazian [13], we use 55% as a threshold for the variance explained by the first principal component.

If this threshold is not achieved, the second, third and other principal components should be taken into construction of the ranking function until the explained variance goes beyond 55%. In formal way, this implies the inequality:

$$\frac{\lambda_1 + \dots + \lambda_k}{\lambda_1 + \dots + \lambda_k + \dots + \lambda_p} \geq 0.55 \quad (4)$$

where λ_l – eigenvalue of the l -th principal component;

k – number of first principal components which should be taken into construction of the ranking function;

p – number of principal components obtained.

Step 6. Specifying the ranking function. Latent variable which reflects the 'worth' and 'quality' of a client is now can be defined as a weighted (by principal components eigenvalues) sum of first k principal component scores

$$\widehat{Score}_i = \sum_{l=1}^k \lambda_l \cdot s_{il} \quad (5)$$

Thus, score weights are being determined by principal component analysis algorithm without a need of expert opinions on the weights.

Step 7. Calculating the values of the ranking function. The function (5) is used to get $Score_i$ for each client ($i \in [1, n]$).

Step 8. Sorting. All clients are sorted by $Score_i$ from maximum to minimum.

V. SYSTEM OF AUTOMATIC ONLINE CLIENTS RANKING

Proposed approach could be implemented within automatic marketing decision-making in an online store. The author's point of view on this issue is presented in Figure 1.

Clients' actions are stored in a database and used for constructing (calculating) a set of variables (characteristic) that indicate 'worth' and 'quality' of a specific client (see Section III). These variables are scaled to the interval $[0, 1]$ and processed by the principal component analysis algorithm. In accordance to the proposed in Section IV, procedure of ranking number of principal components is defined (based on the needed level of explained variance).

Finally, calculation of ranking values is made with subsequent sorting of clients.

In line with the ranking results, a marketing strategy for each client is designed and implemented (e.g. automatically defined amount of discount).

The efficiency of ranking and related marketing strategies should be assessed (e.g. clients loyalty measures over time, stability of ranking results, etc.) and corrective measures, if needed, should be introduced. Strictly saying, such correction is only possible in (i) redefining a set of variables, (ii) changing the needed level of explained variance and (iii) the logic of designing a marketing strategy based on the ranking results. It is an open question whether all of this can be done automatically in principle without intervention of experts.

VI. CONCLUSIONS AND FURTHER WORK

In the article, an approach to unsupervised (without experts and learning sample) clients ranking has been proposed. The approach is based on the known latent variable analysis. Under such a latent variable the ‘worth’ and ‘quality’ of a client is considered.

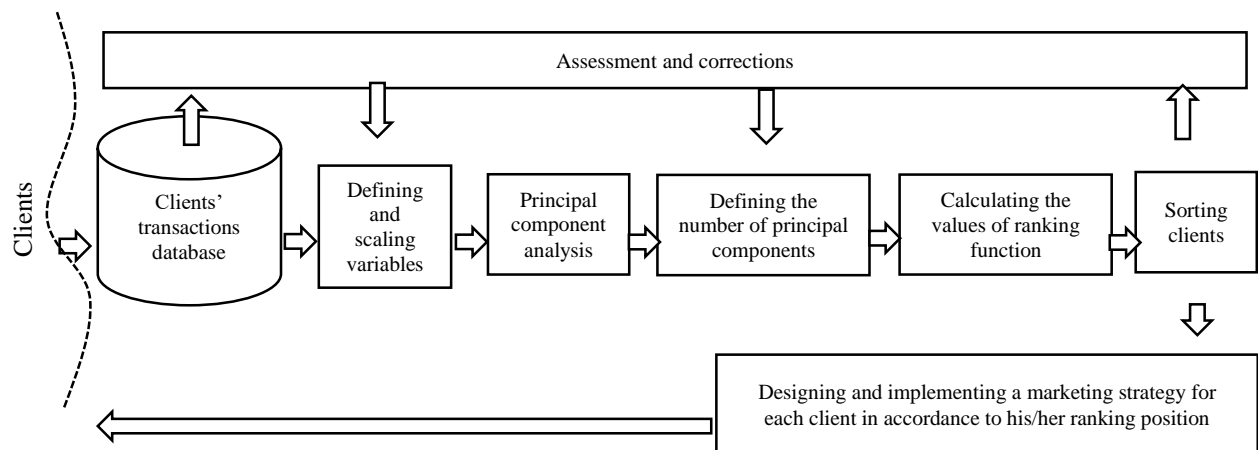


Fig. 1 Conceptual model of the system of automatic clients ranking

Besides this, the conceptual model of the system of automatic clients ranking in the framework of automatic decision-making in online stores has been introduced.

The main question of further research whether it is correct in principle to synthesize an estimate of ‘worth’ and ‘quality’ of client from a set of partial indicators using not expert judgments on real-life importance of indicators, but mathematical properties of the data.

The second question whether ranking based on several principal components (not a single one, the first component) has a business meaning at all, because having two or more principal components means a set of latent variables.

Till now the proposed conceptual model is no more than a concept and it’s still open whether it can be implemented in real context of an online store. As well as it’s not theoretically clear (i) how a marketing strategy can be tailored in accordance to ranking results, (ii) what efficiency measures could be used in the system logic.

Finally, there is an urgent need to test the proposed approach on real datasets.

VII. ACKNOWLEDGEMENTS

The authors thank two anonymous reviewers for their constructive comments, which helped us to improve the manuscript.

REFERENCES

- [1] Klepacki, B., and I. Dziedzic-Jagocka, "Stock Management with the example of the production/logistic company Vive Textile Recykling", *Аграрна економіка*, Volume 3, № 1-2, 2010, pp. 120-130.
- [2] M. Karthick, S. Karthikeyan, and M.C. Pravin, "A Model for Managing and Controlling the Inventory of Stores Items based on ABC Analysis", *Global Journal of Researches in Engineering*, Volume 14, Issue 2, 2014.
- [3] B. Noche, "ABC-/XYZ Analysis Introduction", https://www.uni-due.de/imperia/md/content/tul/download/en_ss2015_lm01_le_abc_analysis.pdf
- [4] Flores, Benito E., and D. Clay Whybark, "Multiple criteria ABC analysis", *International Journal of Operations & Production Management*, 1986, 6(3), pp.38-46.
- [5] Scholz-Reiter, B., Heger, J., Meinecke, C. and Bergmann, J., "Integration of demand forecasts in ABC-XYZ analysis: practical investigation at an industrial company", *International Journal of Productivity and Performance Management*, 61(4), pp.445-451.

- [6] Khajvand, M., Zolfaghar, K., Ashoori, S. and Alizadeh, S., "Estimating customer lifetime value based on RFM analysis of customer purchase behavior: Case study", *Procedia Computer Science*, 2011, 3, pp.57-63.
- [7] Chang, H.C. and Tsai, H.P., "Group RFM analysis as a novel framework to discover better customer consumption behavior", *Expert Systems with Applications*, 2011, 38(12), pp.14499-14513.
- [8] Hosseini, S.M.S., Maleki, A. and Gholamian, M.R., "Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty", *Expert Systems with Applications*, 2010, 37(7), pp. 5259-5264.
- [9] Foster, G., Gupta, M. and Sjoblom, L., "Customer profitability analysis: challenges and new directions", *Journal of Cost Management*, 1997, 10, pp.5-17.
- [10] Berger, P.D. and Nasr, N.I., "Customer lifetime value: Marketing models and applications", *Journal of interactive marketing*, 1998, 12(1), pp.17-30.
- [11] Jain, D. and Singh, S.S., "Customer lifetime value research in marketing: A review and future directions", *Journal of interactive marketing*, 2002, 16(2), pp.34-46.
- [12] Guarino, C., Ridgeway, G., Chun, M. and Buddin, R., "Latent variable analysis: A new approach to university ranking", *Higher Education in Europe*, 2005, 30(2), pp.147-165.
- [13] Aivazian, S., "Synthetic indicators of quality of life: Construction and utilization for social-economic management and comparative analysis", *Austrian Journal of Statistics*, 2005, 34(2), pp.69-77.

Analysis and software implementation of the methods for extracting semantic proximity between the lexical units

S. Leoshchenko, A. Oliinyk, S. Subbotin

Abstract— This paper presents one of the ways solution to the problem of lexical ambiguity: the using methods of extraction and analysis of semantic relations between lexical units. The proposed methods are divided into groups, among which identified and discussed the most popular. As well, special attention is paid to the methods of k-nearest neighbor and mutual k-nearest neighbor, which allow to introduce the problem of extracting semantic proximity as one of the classification tasks.

Keywords—Software tools, semantic relations, semantic proximity, methods and metrics, software system, information extraction, Wikipedia, computational lexical semantics

I. INTRODUCTION

Today artificial intelligence is one of the most popular branches of science and technology. Moreover, every day, even ordinary people, who are not closely linked with science, are increasingly using this phrase. This is simply explained by the fact that more and more people use the device which in varying degrees use the tools and technologies of artificial intelligence.

But even the using of such advanced information technology, often does not allow to solve problems that the ordinary person can decide on a subconscious level. For example, choosing the most correct option from the list of synonyms given the content or the stylistic feature of the context. In such a situation, people in most cases will be able to find the exact variant in less than a second, in contrast to any software product that in turn reduces the efficiency of the use of, for example, automatic translators. One solution to this problem is the determination of semantic proximity. This study focuses on this method of solving a highly relevant problem.

II. ANALYSIS EXISTING METHODS OF EXTRACTING SEMANTIC RELATIONSHIPS

One of the classes of methods of solving the problem of lexical ambiguity are methods based on external knowledge sources. In turn, these methods can be divided into categories, and one category of such methods are methods based on the degree of semantic proximity calculated on the basis of semantic networks. In this category plays an important role completeness and timeliness of the network, and calculated measures of semantic proximity. By itself, the problem of computing semantic proximity resource intensive and requires considerable computing power.

To work with calculation methods of semantic proximity, it is necessary to highlight the existing semantic relations between words in natural languages. Such relationships are: synonyms, meronyms, antonyms, associations, etc.

A. Panchenko, S. Adeykin, A. Romanov and P. Romanov gives examples of successful application of semantic proximity in their works: such relations are successfully used in various NLP applications, such as word sense disambiguation, query expansion, document categorization or question answering [5].

However, existing resources in most cases are unavailable to ordinary users and to evaluate their effectiveness will be possible only according to given scientific articles. Moreover, the creation of the needed semantic resources in the manual would be too time-consuming, lengthy and costly process. This implies that the development and the subsequent creation of methods for automatic extraction and analysis of semantic relations is extremely important.

The most common method of extracting semantic relations based on lexico-syntactic patterns that are created manually [5]. This approach has many disadvantages, chief among which is the complexity of writing rules to

S. Leoshchenko Zaporizhzhya National Technical University, Zaporizhzhya, Ukraine (e-mail: sedrikleo@gmail.com).
A. Oliinyk Zaporizhzhya National Technical University, Zaporizhzhya, Ukraine (e-mail: olejnikaa@gmail.com).
S. Subbotin Zaporizhzhya National Technical University, Zaporizhzhya, Ukraine (e-mail: subbotin.csit@gmail.com).

derive the relations. From this disadvantage, we can conclude that the use of such rules will be correct with respect to only one language, others will have to create new rules. So there are methods based on distribution analysis [6,7]. Their advantage is that they do not require manual work, but the main drawback: low results when used to extract semantic relations [8]. Good results in the tasks demonstrate metrics of retrieve and evaluate semantic proximity, which based in using semantic network in Wikipedia that have been proposed recently [9,10,11]. The relevance of the use of Wikipedia as semantic network analysis is useful because it is one of the largest semantic networks. It covers most of the subject areas, daily updated and supplemented with the users is checked for errors and stores data almost all used languages. And since the existing research Wikipedia was used to test metrics extraction of semantic relations is not often, it can show good results which we had never seen before.

There are different methods based on artificial intelligence [15-20] that can be used for real word processing and analysis. Information technologies using different real word processing methods proposed in [21-25]. But proposed in [15-25] methods and information technologies can't be used for semantic proximity extraction between the lexical units.

A. The content measures

As a rule, the content measures of semantic proximity based on the use of Wikipedia articles as vectors in a space of words or terms. Elements of the vector are usually the frequency of occurrence of relevant terms in the text of the articles or their weight, computed according to the scheme TF-IDF:

$$w_{t,d} = tf_{t,d} \cdot idf_{t,d} \quad (1)$$

where $tf_{t,d}$ is the frequency of occurrence (term frequency) of term t in the text d and

$$idf_{t,d} = \log \frac{|D|}{|\{d \in D \mid t \in d\}|} \quad (2)$$

– inverse term frequency t in megen Wikipedia articles. The semantic proximity of the two articles may be further evaluated as the cosine of the angle between their vector representation.

Method Latent Semantic Analysis (LSA) [13] uses singular value decomposition (SVD-decomposition of the matrix of TF-IDF weights of the terms in the articles to construct the space of the factors (subjects). Further texts can be represented by vectors in the new space to calculate the semantic proximity between them. The development of the LSA method is probabilistic latent semantic analysis (probabilistic LSA pLSA) [14]. Unlike LSA that assumes a normal distribution of documents and terms, pLSA is based on a multinomial distribution and uses a mixed decomposition instead of the SVD decomposition.

One of the drawbacks of LSA is the difficulty of interpretation of the obtained topics. Gabrilovich and Markovich [10] proposed to design the input texts directly on many of the concepts of Wikipedia. Method is called Explicit Semantic Analysis (ESA).

Each Wikipedia concept in the ESA method is represented as a vector of words of the relevant article, weighted according to the scheme TF-IDF. The authors then build an inverted index that displays in the list of articles, where they meet. Concept with a very small weights for a given word are discarded. The input text, denote it by

$T = \{w_i\}$ is represented as a TF-IDF vector of weights of words $\langle v_i \rangle$, where v_i is the weight of w_i and then

displayed in the space of concepts. Let $\langle k_j \rangle$ is the index entry for the key w_i , where k_j is the TF-IDF weight of

the word w_i in the article $c_i, c_j \in c_1, c_2, \dots, c_N$, N is the total number of Wikipedia concepts. Then the

weight of the concept c_j in the vector that matches the text is calculated as $\sum_{w_i \in T} v_i k_i$. The elements of the

received vector reflects semantic similarity of text to appropriate concepts. To calculate the semantic proximity of the two texts, the authors compute the cosine of the angle between the vector representations in the space of concepts.

B. B. Method of the random walk

The random walk model has been successfully used for ranking web pages in Internet search engines, for example, well-known PageRank algorithm [17] based on this model. Olivier in his work [18] carried out a comparative analysis of measures of semantic proximity of Wikipedia articles, based on a model of random walks. He considered several measures based on the measure of green, local PageRank, index cross-referencing and cosine terms of the quality of the ranking of these activities. Grin's measure is determined by the following formula:

$$G_{ij} = \sum_{t=0}^{\infty} \delta_{ij} p^t - v_j. \quad (3)$$

In this formula, P is the stochastic matrix based on the adjacency matrix, where $P_{ij} = \frac{1}{\sum_k A_{ik}}$, v is own stationary vector of the matrix P . This measure corresponds to the time spent wandering by the user in the node j if it is started from node i . Similar to the Grin's measure, as local PageRank is expressed by the following formula:

$$S_{ij} = \left(\sum_{t=0}^{\infty} c(1-c)^t p^t \right)_{ij}. \quad (4)$$

In this formula, the random walk process starts from node i , with equal probability clicks on the links of the graph and with probability c , returns to node i . No factor c , S_i turns into a stationary vector of the Markov chain V .

According to the observations of Olivier, the qualitative results give only the modification of the action of the green, and the local PageRank through a very small diameter turns Wikipedia into a global PageRank, which does not carry information about the semantic proximity. One of the most successful methods was the method of $S_{ij} = G_{ij} \log V_j$; on data obtained from the manual ranking of Wikipedia articles, it gave the best results.

A similar approach is used If, in the proposed method, location, PageSim [19].

On computational efficiency of the method of random walks essentially benefit compared with methods a pair of walks, but at a small diameter Wikipedia the use of these measures means the bypass of the entire reference graph of Wikipedia, as for the problem of counting the proximity of two articles and to rank. That is, the computational complexity of this family of measures for both problems is $O(n)$, where n is the number of links Wikipedia. This efficiency is insufficient for the application of these metrics in the practical development of such technology for multilingual text mining, developed by scientists at RAS – Texterra.

C. Wikipedia Link-based Measure

Approach called WLM (Wikipedia Link-based Measure) was presented in [15]. The authors suggested two local metrics to calculate the semantic proximity between the concepts of Wikipedia.

The first metric considers the set of all articles referenced in the initial concept. Suppose we want to calculate the semantic proximity between the concepts A and B . Denote by W the set of all concepts of Wikipedia, and through T is the set of concepts referenced A or B . Each initial concept $s \in \{A, B\}$ and the $t \in T$ is mapped to the value

$$\text{count}(s \rightarrow t) \cdot \log \frac{|W|}{|\{w \in W \mid w \rightarrow T\}|}, \quad (5)$$

where $\text{count}(s \rightarrow t)$ is the number of links from s to t , $\{w \in W \mid w \rightarrow T\}$ is the set of all articles that link to t . It is easy to see the similarity of this scheme with TF-IDF, but with a difference: instead of the terms of reference are resolved. Semantic proximity of concepts A and B is then computed as the cosine between the $|T|$ -dimensional vectors of weights.

The second metric is based on Google normalized distance (Normalized Google Distance) [16], based in turn, on the count of occurrences of terms that represent the desired concepts in the search query results in Google. Web pages containing both terms signal the presence of semantic links between them. WLM version operates instead of the query results links articles:

$$\text{sim}(A, B) = \frac{\log(\max(|\text{in}(A)|, |\text{in}(B)|)) - \log(|\text{in}(A) \cap \text{in}(B)|)}{\log(|W|) - \log(\min(|\text{in}(A)|, |\text{in}(B)|))} \quad (6)$$

where A, B are concepts, issues, $|in(A)|, |in(B)|$ – a lot of articles that reference A and B respectively, W is the set of all concepts of Wikipedia.

III. A COMPARISON OF EXISTING METHODS

The study of existing computing methods of semantic proximity of concepts based on Wikipedia was discovered unsatisfactory outputs, use the values of the lengths of the shortest paths between concepts in the Wikipedia graph, compared with the state-of-the-art methods (which include themed events and activities on the basis of the random walk). However, it has been suggested that not all types of links contribute equally to the semantic closeness between the concepts, and the more accurate their ranking, which other researchers had not previously carried out, can improve the quality are determined with the help of their estimated location.

Also the study revealed the lack of standard datasets for testing measures of proximity between the concepts of Wikipedia, because the vast majority of methods working with concepts not directly but through the terms or texts. For the latter, in turn, there are standard datasets, and their adaptation to the problem of estimating the semantic proximity of the concepts can have on power in further studies.

IV. STATEMENT OF THE PROBLEM

The main goal of a software product that will be developed can be represented as follows: for each word C_i from the input set of words $C = \{C_1, C_2, C_3, \dots, C_n\}$ to find pairs of semantically related words, i.e.

$$R = \left\{ \langle C_i, C_j \rangle, \dots, \langle C_i, C_k \rangle \mid C_i \neq C_j \neq C_k, i, j, k \leq n \right\}.$$

It should also be noted that the methods to be used, do not return the type of the found context, i.e., $R \subseteq C \times C$. Techniques characterized by efficiency, suitability for use with languages available in Wikipedia and sufficient accuracy. The novelty of this work compared to existing research and development is as follows: proposed and investigated and explored new methods for extracting semantic relations from Wikipedia articles based on the algorithms of nearest and mutual nearest neighbors and the two metrics semantic proximity of the words (the cosine of the angle between vectors definitions and the total Lemma in the definitions).

V. ANALYSIS OF METHODS AND METRICS THAT WILL BE USED

A. K-nearest neighbor method

K-nearest neighbor method is a simple non-parametric classification method where a classification of objects within the space of properties use distance (Euclidean, generally), is calculated to all other sites. Select the objects to which the smallest distance, and they are allocated in a separate class.

Method k-nearest neighbors is a metric algorithm for automatic classification of objects. The basic principle of nearest neighbor is that the object is assigned the class which is the most common among the neighbors of this element. The neighbors are taken on the basis of a set of objects whose classes are already known, and, on the basis of the key for the method k values, is calculated, which class is the most numerous among them. Each object has a finite number of attributes (dimensions). It is assumed that there is a certain set of objects with the existing classification [26].

In the most general form, the algorithm of nearest neighbors is:

$$a(x) = \arg \max_{y \in Y} \sum_{i=1}^m [x_{i,x} = y] \omega(i; x), \quad (7)$$

where $\omega(i; x)$ is a given weight function, which evaluates the degree of importance of the i -th neighbor for the classification of the object u . So, if $\omega(i; x) = 1$ for $i < k$, the algorithm corresponds to method k-nearest neighbors. But the problem with several possible answers, the maximum amount of votes can be achieved in several classes simultaneously. The ambiguity can be eliminated if the weight function to take a non-linear sequence, such as a geometric progression: in this example, $\omega(i; x) = [i \leq k] q^i$, which corresponds to the method of exponentially weighted k nearest neighbors, and assume $0.5 \leq q \leq 1$ [27].

Among the main advantages of this method are:

- the simplicity of the implementation;
- the classification, carried out the algorithm, it is easy to interpret by presenting to the user several objects.

Among the main disadvantages of this method are:

- excessive complexity of decision rule because of the need of storing the training sample;

- the nearest neighbor search involves the comparison of an object is classified, together with all objects of the sample, which requires linear in the length of the sampling operations.

B. Mutual k -nearest neighbor method

The method of mutual nearest neighbors is a method to consider mutually k -nearest neighboring data points, not just the nearest neighbor [28].

Let $N_k(x)$ is the set of k nearest neighbors of x in D_n , $N'_k(x_i)$ the set of k nearest neighbors x_i in $(D_n \setminus \{x_i\}) \cup \{x\}$. A set of mutual k -nearest neighbor (MkNNs) x is defined as:

$$M_k(x) = \{x_i \in N_k(x) : x \in N'_k(x_i)\} \quad (8)$$

Then the mutual k -nearest neighbor is defined as:

$$m_n^{MkNNR}(x) = \begin{cases} \frac{1}{M_k(x)} \sum_{i: x_i \in M_k(x)}^k y_i, & \text{if } M_k(x) \neq 0 \\ 0, & \text{if } M_k(x) = 0 \end{cases} \quad (9)$$

whereas $M_k(x) = |M_k(x)|$.

In General, if we compare the method of mutual nearest neighbors, with a limit based on the mutual KNN, it should be noted that Mutual KNN better able to identify clusters of various shapes and sizes. Moreover, this method is less prone to noise in the data and can detect deviations [29].

C. Metric number of the common Lemma in the word definitions

The metric uses a measure of semantic proximity based on shared words in the definitions of the two terms.

$$similarity(t_i, t_j) = \frac{2|(d_i \cap d_j) / stopwords|}{|d_i| + |d_j|} \quad (10)$$

The numerator of the fraction equals the total number of words in the initial form, without the list of words defined as stop words.

The denominator is the sum of all words in each of the two chosen definitions.

However, this metric does not take into account the length of the definitions. This is the main drawback of the metric, since the length of some definitions can reach up to one hundred rows and have a large number of commonly used terms that are not included in the list of stop words, for example: system, population, range, amount etc [30]

D. Metric the cosine of the angle between the definitions

In order to compensate for the effect of the length of definitions in the connectivity between the terms, used a metric called "the cosine of the angle between vectors definitions". The definition is represented as an N -dimensional vector, and then evaluates the time using the obtained vectors.

The dimension of the vector is determined depending on the nature of the investigated problem. Thus, the vector definition will include only terms from the dictionary. If the dictionary is not found in the definitions, we denote the element becomes zero, in other cases, the element becomes equal to the number of occurrences in the definition of the term.

$$similarity(t_i, t_j) = \frac{f_i \cdot f_j}{\|f_i\| \cdot \|f_j\|} = \frac{\sum_{k=1, N} f_{ik} f_{jk}}{\sqrt{\sum_{k=1, N} f_{ik}^2} \sqrt{\sum_{k=1, N} f_{jk}^2}} \quad (11)$$

In the formula of cosine similarity between the numerator represents the dot product of the two above definitions, and the denominator is equal to the product of the Euclidean norms of these vectors. The denominator in the formula normalizes by the length of the vectors, so the result can be interpreted as the scalar product of normalized vectors corresponding to the two definition.

To implement the metrics necessary to pre-treatments:

- the normalization of definition – bringing the definition of each word in the initial form;
- normalization dictionary – bringing each word dictionary in primary forms;
- the search terms used in the definition (the search takes into account all words in terms of the dictionary);
- the formation of vectors definition of that dimension, which would equal the number of terms in the dictionary [30].

VI. THE ALGORITHM OF THE SOFTWARE PACKAGE

Methods of extraction of semantic relations, which are used based on component analysis [35, 36], the principle of which is that semantically similar words have similar definitions. The proposed algorithms use one of two metrics of similarity denitions, the number of common words [37] or the cosine of the angle between vectors definitions [38]. As input algorithms for extracting semantic relationships take a lot of words C , between which it is necessary to calculate the ratios and their definitions D .

The first version of the algorithm computes semantic relations using the method of nearest neighbors KNN, the second – using the method of mutual nearest neighbors MKNN (Mutual KNN). The only metaparameter algorithms the number of nearest neighbors k . The pseudocode of the algorithms is shown in Fig. 1.

```

1.  // Calculation of pairwise similarities between words all concepts C
2.  Rmatrix = void
3.  for i=0; i<count(C); i++ {
4.      for j=i; j<count(C); j++ {
5.          // Calculation of semantic similarity of two concepts
6.          s_ij = similarity (D(i) , D(j))
7.          // Saving most similar concepts
8.          if( count(Rmatrix(C(i))) < k || s_ij > min(Rmatrix(C(i))) ){
9.              Rmatrix(C(i)).addOrReplaceMin(C(j))
10.         }
11.     }
12. }
13. // Calculation of semantic relations
14. R = void
15. foreach c_i in Rmatrix {
16.     foreach c_j in Rmatrix(c_i) {
17.         if(!isMutualKNN || Rmatrix(c_j) contains c_i){
18.             R.add(<c_i, c_j>)
19.         }
20.     }
21. }
22. return R

```

Fig.1 Pseudocode of the algorithms

Algorithms is the following. First, calculate a measure of semantic proximity of all possible pairs of definitions (line 6). On the basis of the calculated values filled in the array closest words Rmatrix for each definition (lines 1-12). The number of array elements supported is identical to k (number of nearest neighbors) – this allows to strongly reduce memory consumption without loss of information about the connectivity of the words. After filling the array the most similar words for each definition all that remains to be done to retrieve the result set of relations R in KNN method – just fill in the source set, and for the MKNN method is to check for each definition: it is an array of the most similar words pair, and if included – add to the result set (lines 13-21).

The complexity of the algorithms is proportional to the amount fed to the input of the words $|C|$. Time complexity is $O(|C|^2)$, space complexity is also proportional to the number of nearest neighbors k is $O(k|C|)$.

VII. EXPERIMENTS AND RESULTS

Investigated the algorithms KNN and MKNN with the two above-described metrics of closeness and different values of the number of nearest neighbors k , the results clearly demonstrated using the diagram in Fig. 2. The results indicate almost linear increase in the number of relationships detected, depending on the parameter k for both algorithms. The number of found relations is only slightly dependent on the used metric. The KNN algorithm gets more relations than MKNN, with an equal number of nearest neighbors k . This is because MKNN removes pairs that are not mutual neighbors, in contrast to the KNN.

Also evaluated the accuracy of algorithms KNN and MKNN for $k = 2$ of the set with 775 definitions. For this was the marked files manually to remote relationships and calculated the precision of the retrieval as the number of true relations to the amount of the withdrawn relationship. The results are shown in the table 1.

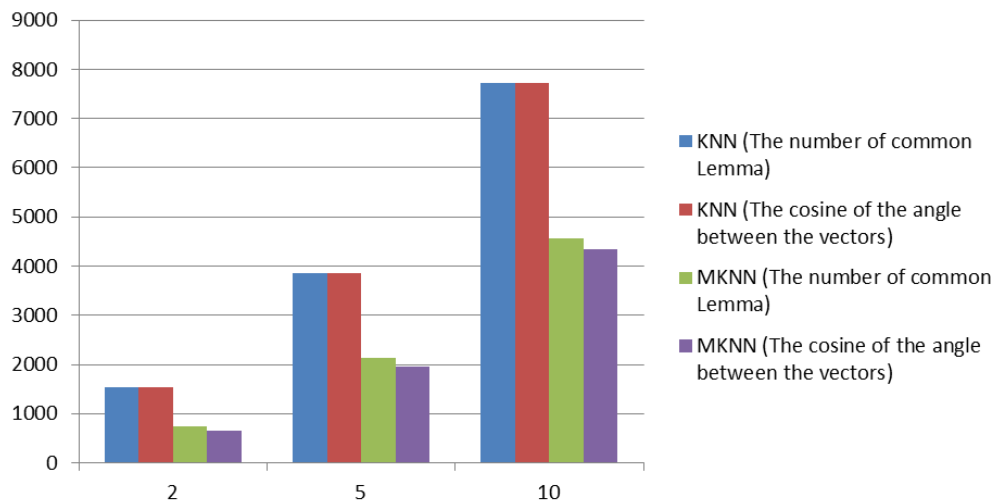


Fig.2 Results

Table 1 The accuracy of extraction methods KNN and MKNN if $k = 2$

Method	The Metric	Is Taken	Correct	Accuracy
KNN	The number of common Lemma	1546	1167	0.761
	The cosine of the angle between the vectors	1546	1167	0.761
MKNN	The number of common Lemma	724	603	0.833
	The cosine of the angle between the vectors	652	499	0.763

Due to the large number of recovered relationships (Fig. 2), evaluation of extraction manually for all values of k is heavy. For large values of k the accuracy of the extraction of the relations needs to decrease. When using the method I recommend using $k \in [1;10]$. In the future I plan to use WordNet and a standard testing set of semantic relations, such as BLESS [39], for a more accurate evaluation of the quality of retrieval.

VIII. CONCLUSION

The article considers new methods for extracting semantic relations. Despite the fact that this are simple methods for data mining, they demonstrate very good results. To assess the quality of methods used universal metrics which complement each other. The simplicity of the proposed methods and metrics ensures stable operation and high performance of the developed software.

Developed a software package "SemAnalysis", implements the basic functions: analysis and extraction of semantic relations, namely semantic proximity between words. This program uses methods: nearest neighbors and mutual nearest neighbors and metrics: the total number of LEM in the definitions and the cosine of the angle between vectors definitions. Also the software package uses a large data base of terms and their definitions.

Due to the fact that the program is created using C++ programming language, it can be running on the computer running Windows or UNIX-like operating system.

Developed the software package has a practical focus and is ready to use.

During the development of the program and analysis of the developed program complex has some ideas for improvement such as:

- the use of modern document-oriented database system – MongoDB;
- introduction parallel programming algorithm of the program.
- continued use of popular semantic networks – WordNet and Wikipedia.

New opportunities will be able to expand the functionality of the software and enhance its performance.

ACKNOWLEDGMENT

This article is the result of the project implementation: TEMPUS CERES Centers of Excellence for young REsearchers, reg.no. 544137-TEMPUS-1-2013-1-SK-TEMPUS-JPHES

REFERENCES

- [1] Patwardhan S., Pedersen T. Using WordNet-based context vectors to estimate the semantic relatedness of concepts. EACL, 2006. P. 1-12.
- [2] Hsu M.H., Tsai M.F., Chen H.H. Query expansion with conceptnet and wordnet: An intrinsic comparison. Information Retrieval Technology. LNCS, 2006. P. 1–13.
- [3] Tikk D., Yang J.D., Bang S.L. Hierarchical text categorization using fuzzy relational thesaurus. Prague: KYBERNETIKA, 2003. P. 583–600.
- [4] Sun R., Jiang J., Fan Y. Using syntactic and semantic relation analysis in question answering. In Proceedings of the. TREC, 2005. 243 p.
- [5] Panchenko A., Adeykin S., Romanov A., Romanov P., Extraction of Semantic Relations between Concepts with KNN Algorithms on Wikipedia. URL: <http://it-claim.ru/Projects/RFH/Publications/2%20kv/2012%2022%202%20Panchenko.pdf>
- [6] Hearst M.A., Automatic acquisition of hyponyms from large text corpora, Proceedings of the 14th conference on Computational linguistics. COLING, 1992. P. 12-18.
- [7] Lin D. Automatic retrieval and clustering of similar words. Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics. Montreal: COLING-ACL, 1998. P. 768-774.
- [8] Heylen K., Peirsman Y., Geeraerts D., Speelman D. Modelling word similarity: an evaluation of automatic synonymy extraction algorithms. Proceedings of the Sixth International Language Resources and Evaluation. Leuven: LREC, 2008. P. 3243-3249.
- [9] Curran J.R., Moens M. Improvements in automatic thesaurus extraction. Proceedings of the ACL-02 workshop on Unsupervised lexical acquisition. ACL, 2002. P. 59-66.
- [10] Strube M., Ponzetto S.P. WikiRelate! Computing semantic related-ness using Wikipedia. Proceedings of the National Conference on Artificial Intelligence. Heidelberg: AAAI Press, 2006. P. 1419-1429.
- [11] Gabrilovich E., Markovitch S. Computing semantic relatedness using wikipedia-based explicit semantic analysis. International Joint Conference on Artificial Intelligence. Israel: Israel Institute of Technology, 2007. P. 12-20.
- [12] Zesch T., Müller C., Gurevych I. Extracting lexical semantic knowledge from wikipedia and wiktionary. Marrakech: LREC, 2008. P. 1646–1652.
- [13] Resnik P. Using information content to evaluate semantic similarity in a taxonomy. San Francisco: Morgan Kauffman Publishers Inc., 1995. P. 448–453.
- [14] Deerwester S. C. et al. Indexing by latent semantic analysis. Tokio: JASIS, 1990. P. 391-407.
- [15] Hofmann T. Probabilistic latent semantic analysis. San Francisco: Morgan Kaufmann Publishers Inc., 1999. P. 289-296.
- [16] Witten I., Milne D. An effective, low-cost measure of semantic relatedness obtained from Wikipedia links. Chicago: AAAI Press, 2008. P. 25-30.
- [17] Cilibrasi R. L., Vitanyi P. M. B. The google similarity distance. Amsterdam: IEEE Transaction, 2007. P. 370-383.
- [18] Lee M. D., Pincombe B. M., Welsh M. B. An empirical evaluation of models of text document similarity. Cognitive Science, 2005. P. 12-18.
- [19] Velikhov P.E. Measures of semantic proximity of Wikipedia entries and their application at text processing. SPb.: Technologies and Computer Systems, 2009. P. 23-37. (In Russian).
- [20] Akiba T., Iwata Y., Yoshida Y. Fast exact shortest-path distance queries on large net-works by pruned landmark labeling. New York: ACM, 2013. P. 349-360.
- [21] Székely G. J. et al. Brownian distance covariance //The annals of applied statistics. Annals of Applied Statistics, 2009. P. 1236-1265.
- [22] Bellet A., Habrard A., Sebban M. A Survey on Metric Learning for Feature Vectors and Structured Data. arXiv preprint arXiv, 2013. 59 P.
- [23] Xing E. P. et al. Distance metric learning with application to clustering with side-information. California: NIPS, 2003. P. 521-528.
- [24] Senellart P., Blondel V. D. Automatic Discovery of SimilarWords. Survey of Text Mining II. London: Springer, 2008. P. 25-44.
- [25] WordNet. URL: <https://wordnet.princeton.edu/>.
- [26] Textterra. URL: http://www.ispras.ru/proceedings/isp_26_2014_1/isp_26_2014_1_421/.
- [27] K Nearest Neighbor Algorithm. URL: http://www.d.umn.edu/~deoka001/downloads/K_Nearest_Neighbor_Algorithm.pdf.
- [28] Introduction to k-nearest neighbors : Simplified. URL: <https://www.analyticsvidhya.com/blog/2014/10/introduction-k-neighbours-algorithm-clustering/>.

- [29] Kim Hyun-Chul Bayesian Kernel and Mutual k-Nearest Neighbor Regression. Beijing: IGARSS 2016. P. 32-40.
- [30] Ruan J. A Fully Automated Method for Discovering Community Structures in High Dimensional Data. Department of Computer Science. San Antonio: One UTSA Circle, 2007. P. 7-15.
- [31] Karyaveva M.S. Linguistic and Statistical Analysis of the Terminology for Constructing the Thesaurus of a Specified Field. Modeling and Analysis of Information Systems. Jaroslavl': JaGU, 2015. P. 18-36. (In Russian).

FPGA LUTs for a Logic Systems

Tyurin Sergey, Prokhorov Andrey, Vikhorev Ruslan

Abstract— FPGA logic based on Look up Tables (LUTs). However, LUT calculates only one logic function in the perfect disjunctive canonical forms (PDCF) for this configuration. The paper proposed the concept of the logic advanced LUTs in three main areas. The first area is Double LUT (DLUT), computes two functions simultaneously with inactive transmission transistors subtree. The second area is DC LUT through inverse this tree to implement the decoder DC, which allows computes whole system of the logic functions. Such technique can significantly reduce hardware expenses for logic systems. The third area is DNF-LUT, which allows the calculations of the system functions in disjunctive normal form (DNF) and else more significantly reduces large number of variables LUTs hardware costs. We explored the models of the proposed concepts in the NI Multisim 10 by National Instruments Electronics Workbench Group. The paper analyzes the assessments of the complexity of the LUT, the conclusions about the effectiveness of the proposed solutions.

Keywords— look up table; FPGA; perfect disjunctive canonical forms.

I. INTRODUCTION

FPGA chips are widely used in computer technology [1-3]. There are quite a large number of energy-saving methods of configuring FPGA [1, 2] for example, such as energy-efficient mapping and clocking, unused blocks power down and others. In Ph.D. Dissertation [4] suggested an FPGA post-fabrication component-specific mapping and an optimized architecture taking into account the characteristics of individual transistors, identified during the operational phase. It uses minimum energy/operation indicator. However, the expansion of the actual FPGA logic capacity for energy efficiency presented in the available sources not full. For this, it is possible to use logic optimization [5-7]. This particularly applies to the implementation of logical systems, an example of which given by CPLD [8].

A goal of the paper is research and development logic elements-LUTs by reducing the complexity of the realization of logic functions of a large number of arguments. It presented LUT PDCF technique – DLUT & DC LUT, computes two and more functions in perfect disjunctive canonical forms (PDCF) simultaneously. Secondly is devoted LUT DNF technique by analogy with Programmable Logic Array (PLA). Described the comparison of the number of transistors to implement system of the logic functions in the known LUT and in all proposed LUT.

II. METHODOLOGY

A. LUT FPGA Calculates Two Functions Simultaneously

Consider the LUT model on the two variables (2-LUT), configured to calculate the exclusive OR function is shown in Fig. 1.

S. F. Tyurin, Perm National Research Polytechnic University, Perm, Russia (e-mail: tyurinsergfe@yandex.ru).

A. S. Prokhorov, Perm National Research Polytechnic University, Perm, Russia (e-mail: npoxop007@yandex.ru).

R. V. Vikhorev, Perm National Research Polytechnic University, Perm, Russia Japan (e-mail: vihrusvla@mail.ru).

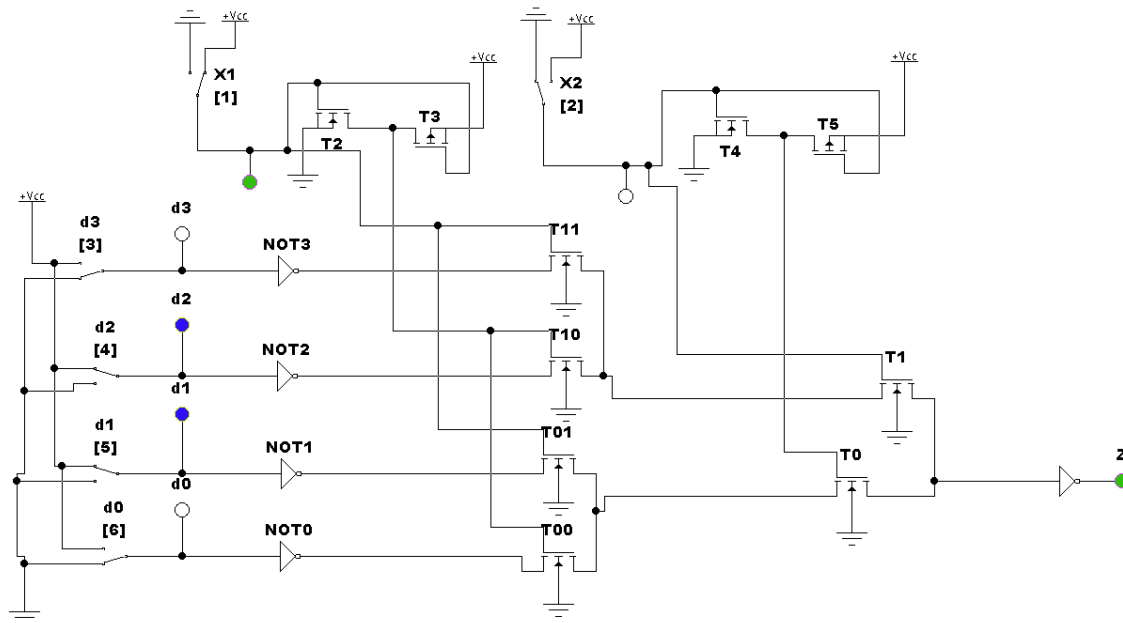


Figure 1. The 2-LUT model configured to calculate the exclusive OR function.

LUT is allocated only one logic function z , customizable by the user by downloading the configuration memory SRAM (d0-d3). At the same time during the computation is always activated only half the tree transistors (T0&T00 or T01 when $x_2 = 0$; T1&T10 or T11 when $x_2 = 1$). This creates the conditions for the use of idle half of the transistors with the introduction of another pair of leading variable. However, this requires connectivity configuration memory SRAM, which stores the settings of the second function. The corresponding truth table is stored "backwards" compared to the truth table of the first function, Fig. 2.

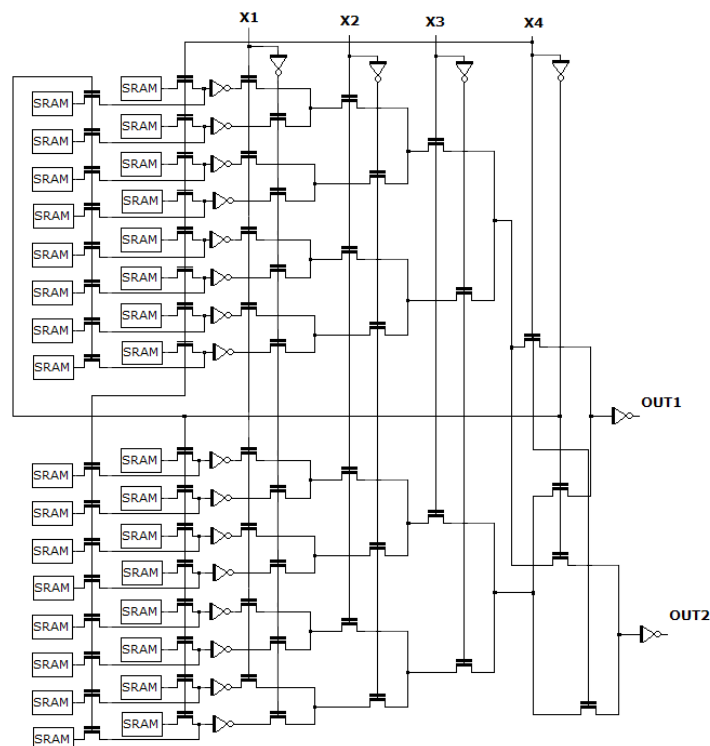


Figure 2. Double4-LUT (D4-LUT) calculates two functions simultaneously

The simulation of the Double2-LUT (D2-LUT) calculates two function $z_1 = x_1 \leftrightarrow x_2$, $z_2 = x_1 \oplus x_2$, $x_1 = x_2 = 0$ in the system NI Multisim 10 is presented in Fig. 3-5. To use the second half of the tree transistors introduced additional transistors leading variable T0.1 and T1.1 and transistors connection settings first function, T00.1, T01.1, T10.1, T11.1, second function T00.2, T01.2, T10.2, T11.2. The setting will simulate constant connection - supply pins V_{cc} and Ground - "zero volts". Simulation confirms the efficiency of the proposed scheme DLUT.

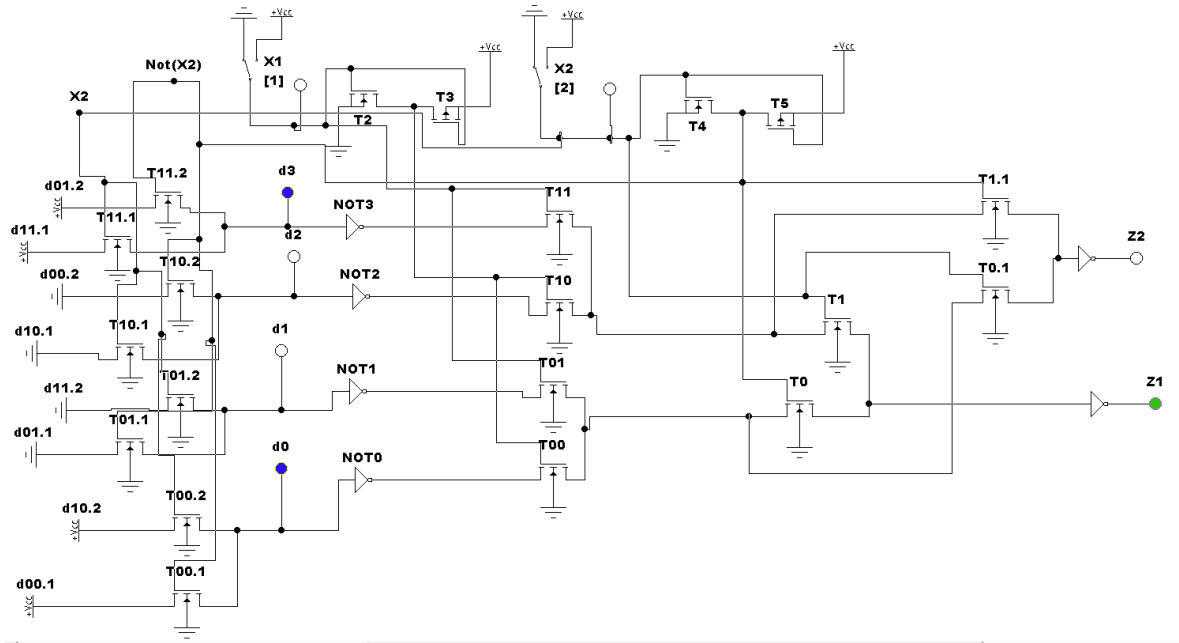


Figure 3. The simulation of the D2-LUT

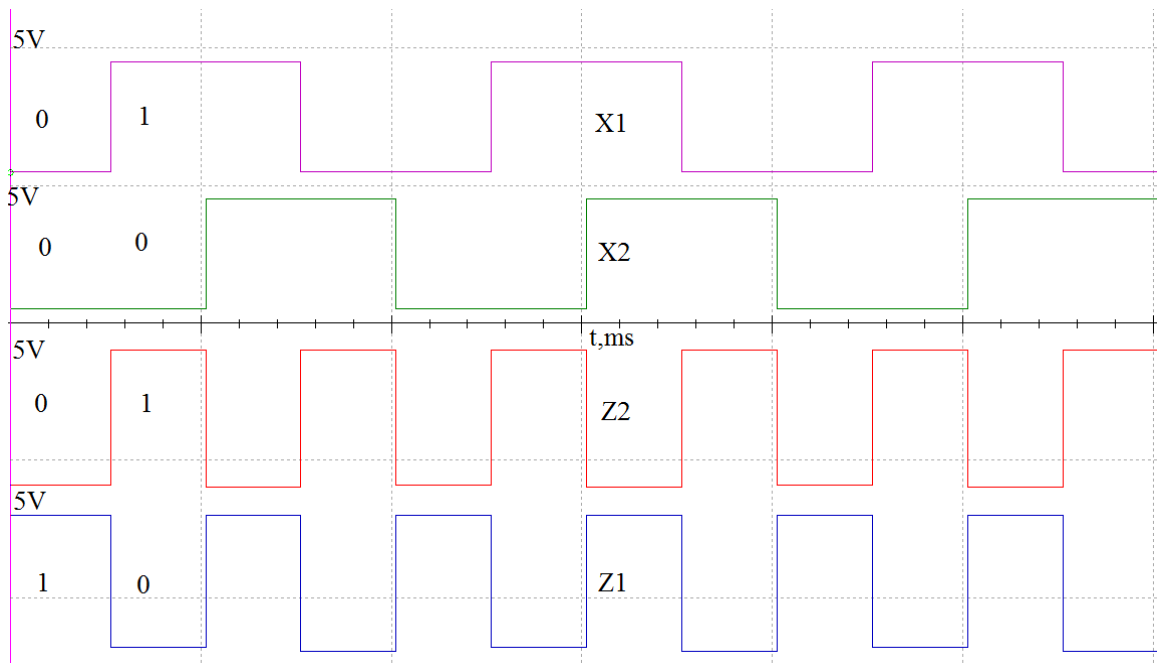


Figure 4. D2-LUT – waveform

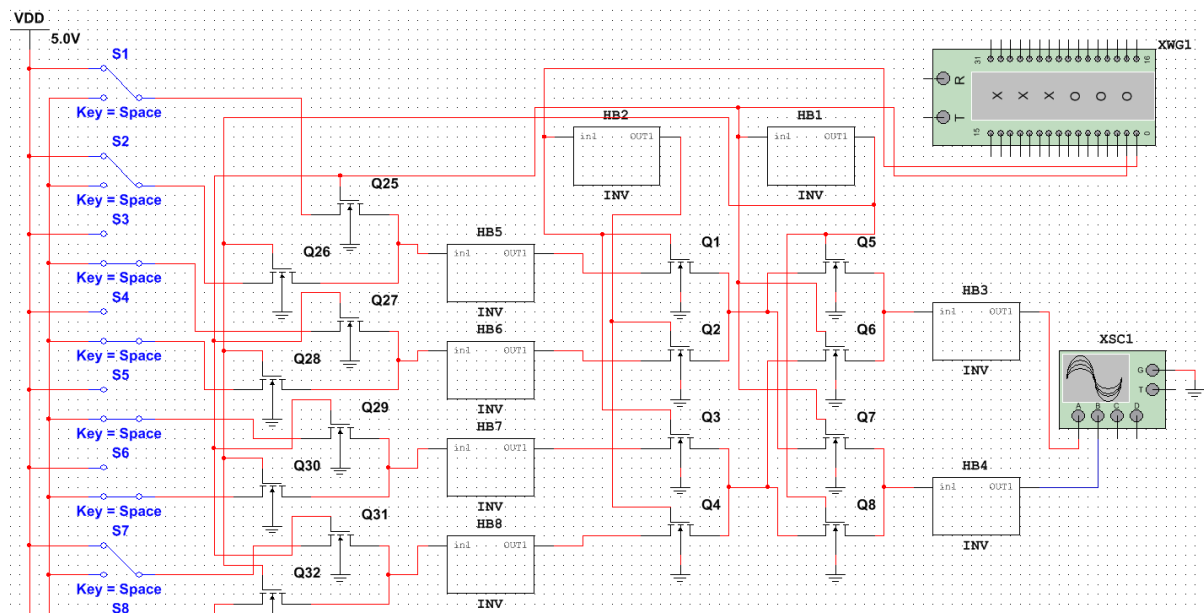


Figure 5. D2-LUT - dynamic model

B. D C LUT FPGA, Calculates System of the Logic Function

The transistor tree "reverse" structure LUT (see. Fig. 1) obtained by "reflection" on the LUT horizontally [9], Fig. 6.

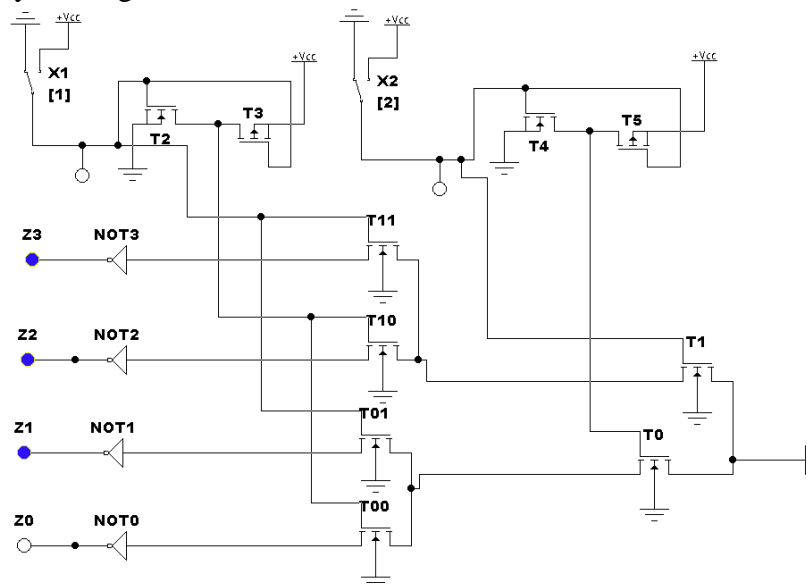


Figure 6. "Reverse" 2-LUT structure

In compliance with the design rules circuits of the transmission transistors required for a drain of each transistor T00, T01, T10, T11 (Fig. 14) to create an alternative chain, transforming its output is guaranteed, for example, in a logic "1". The best option is to create an alternative transistor for each transmission transistor - as it presented in Fig. 7.

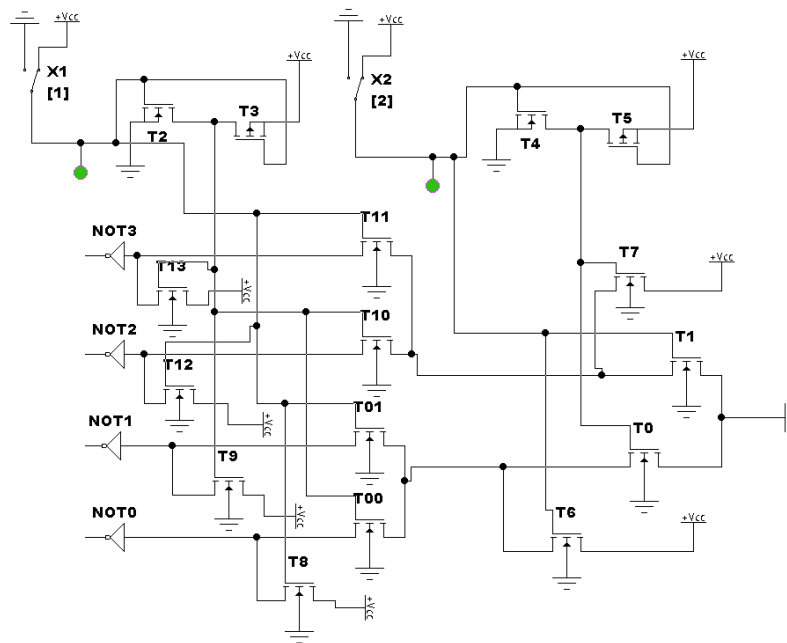


Figure 7. Reverse tree 2-LUT (DC 2-LUT) with alternate transistors T6, 7, 8, 9, 12, 13.

Get the decoder with the output function z (without alternative chains). Further, m times by combining the OR, the corresponding outputs of the outputs we get the implementation of the system of m n -bit logic functions based on perfect disjunctive normal forms (PDNF). DC 2-LUT with one programmable unit disjunctions configured to implement $x_1 \oplus x_2$ show Fig. 8-10.

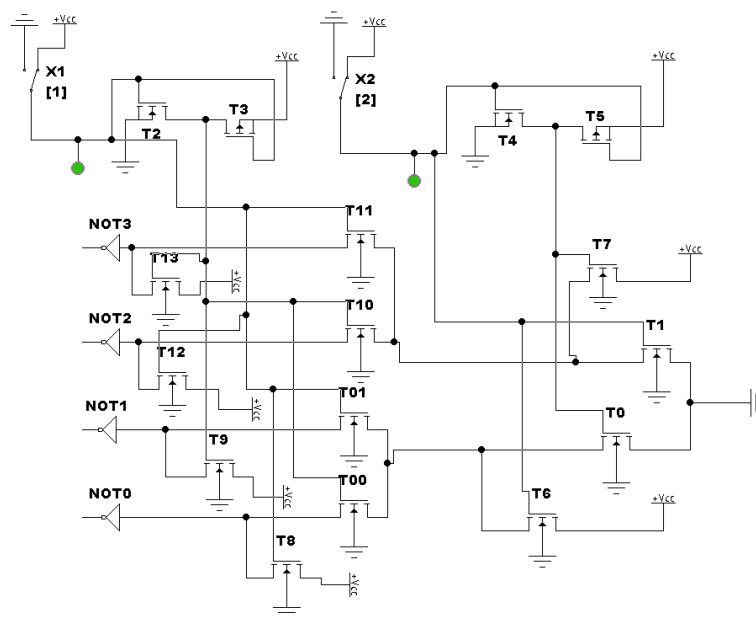


Figure 8. DC 2-LUT - static model

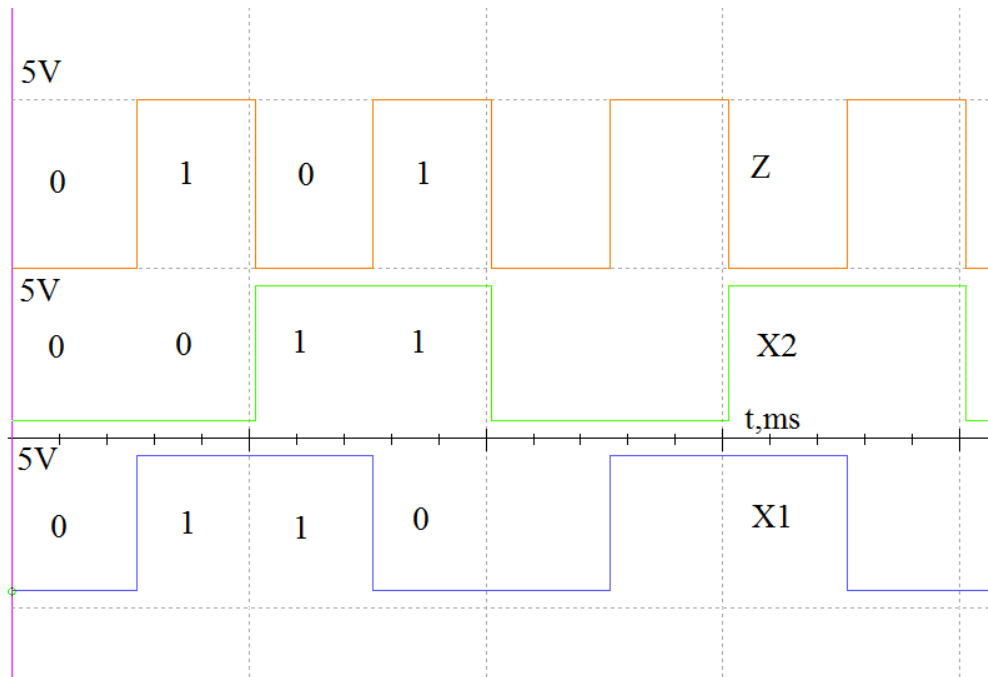


Figure 9. DC 2-LUT – waveform

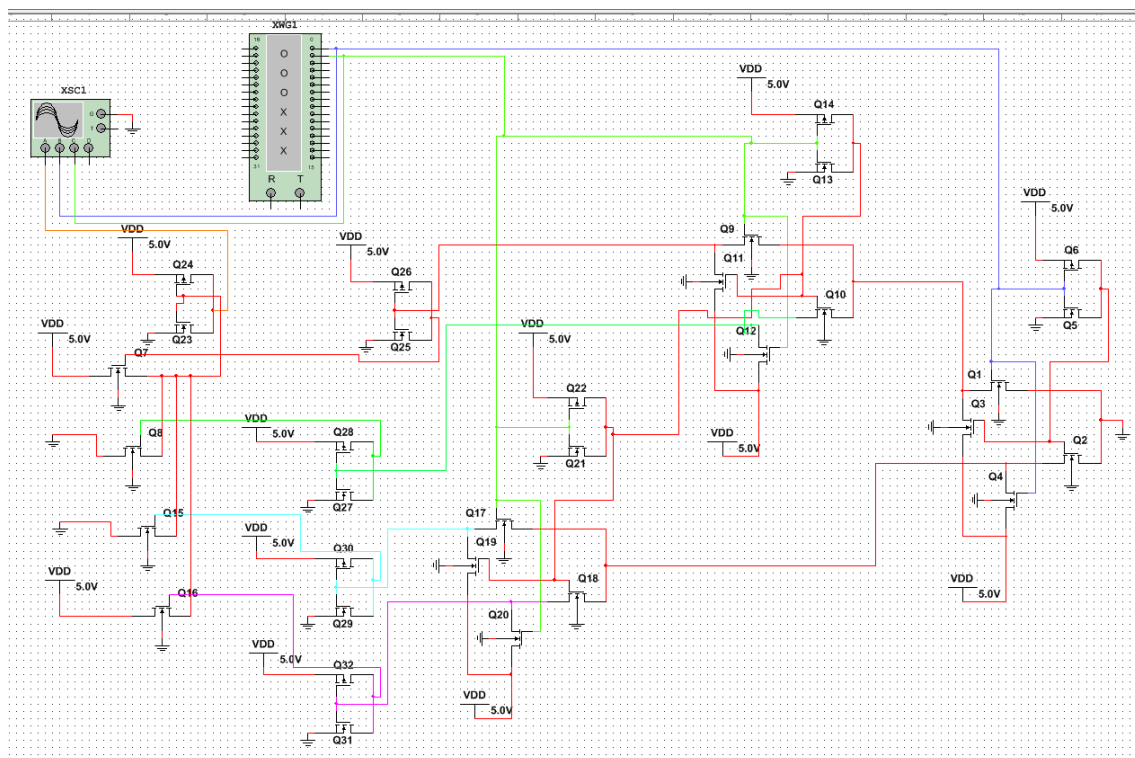


Figure 10. DC 2-LUT - dynamic model

C. Advanced LUT FPGA for Disjunctive Normal Form (DNF) of the Logic Functions Architecture DNF -LUT

The proposed new DNF-LUT [10] is a user-configurable structure similar to a programmable logic array PLA. Architecture DNF-LUT presented on the Fig. 11.

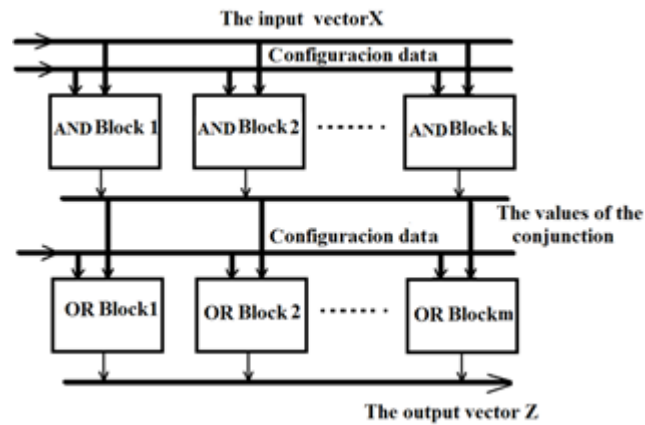


Figure 11. Architecture DNF-LUT

Thus, instead of loading the truth table, it is loading only values programmed conjunctions length n , where n - the number of variables of m logic functions. Occurrences k conjunctions in m functions are also programmable tuning functions. For a given input set (vector n variables x) k AND blocks calculate value of k conjunctions, which then form "an OR" value of m logic functions. The proposed structure of the AND DNF-LUT block is shown in Fig. 12.

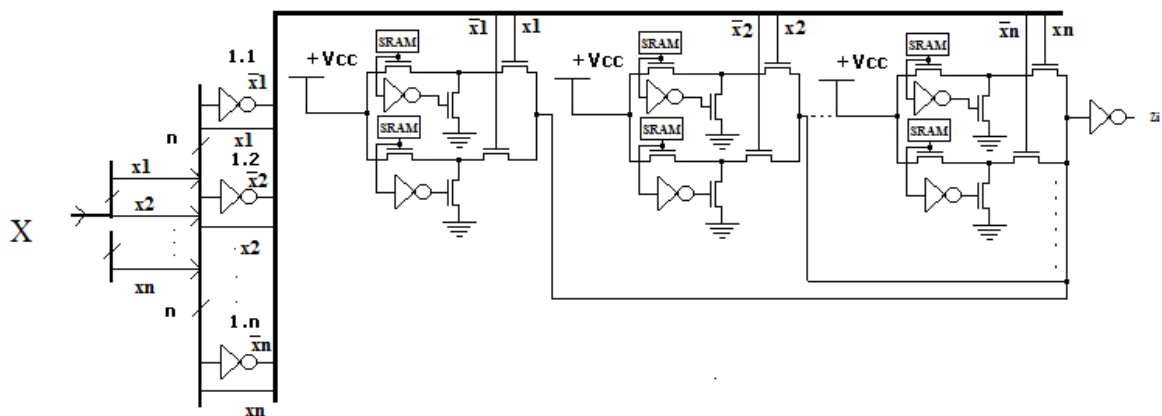


Figure 12. Architecture of the AND DNF-LUT block

One variable SRAM Setup is determined as shown in Fig. 13.

SRAM X	SRAM not X	Output 1	Output 0
1	0	When X	When not X
0	1	When not X	When X
1	1	Anyway	-
0	0	Banned	Banned

Figure 13. One variable AND DNF-LUT SRAM Setup

Thus, if the variable activated, "right", block AND, transmits a logic one signal from the input (left) to the output (right). The same occurs with the immateriality of the variable, i.e., for any value of the variable. If activated "wrong" variable, using inverters and additional transmission transistors supply to the output logical zero. If all variables are the "right" - the output z_i is a logical zero. The proposed structure of the block OR DNF-LUT show Fig. 14.

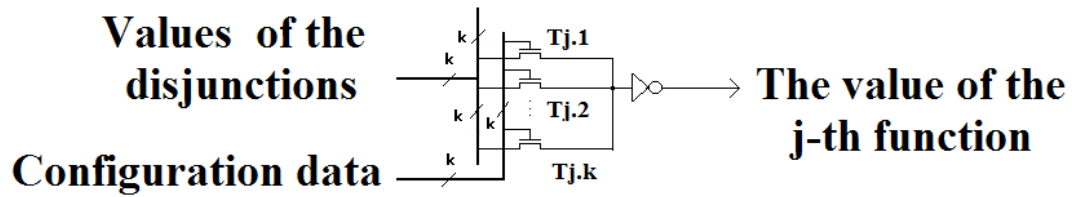
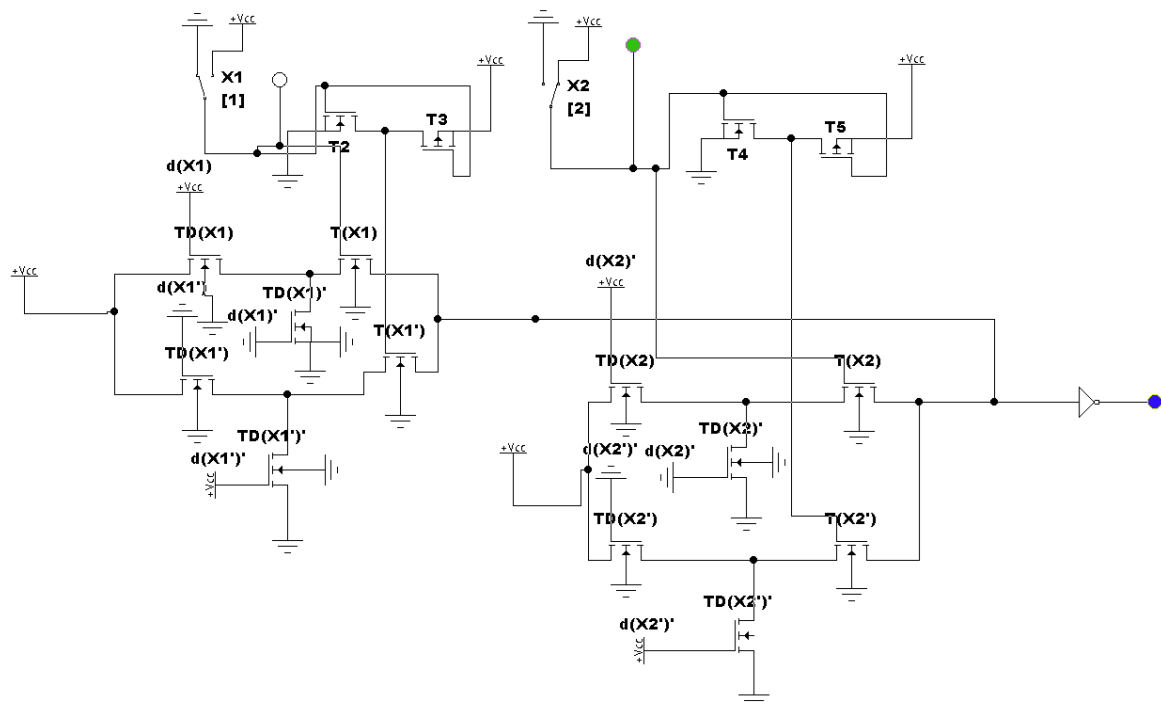


Figure 14. Architecture OR DNF-LUT block

A logical “1” at the output of the corresponding function activated when the inputs given conjunctions are zeros. The simulation of the block AND DNF-LUT executed in the system NI Multisim 10 by National Instruments Electronics Workbench Group show Fig. 15.

Figure 15. AND DNF-LUT Multisim model: $x_2x_1 = 0$, since $x_2 = 1, x_1 = 0$

Given a conjunction $x_2x_1 = 0$, since $x_2 = 1, x_1 = 0$. In case $x_2 = 1, x_1 = 1$ a conjunction $x_2x_1 = 1$ - Fig. 16.

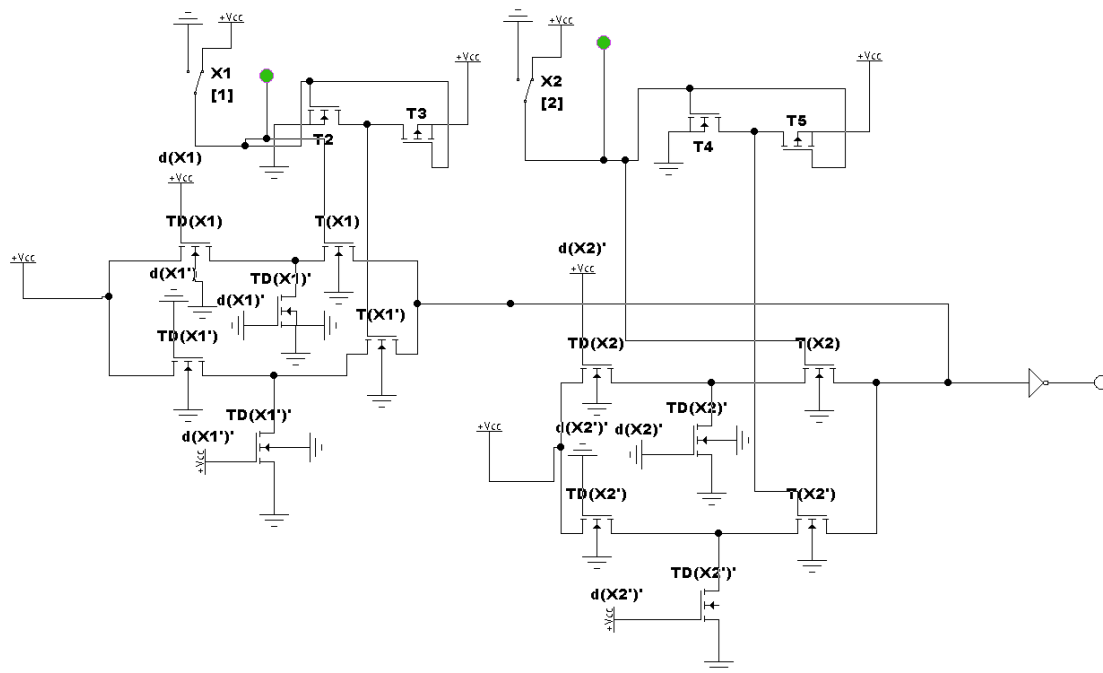


Figure 16. AND DNF-LUT Multisim model: $x_2 x_1 = 1$, since $x_2 = 1, x_1 = 1$

Checking the operation of the rest of the sets also confirms the efficiency of the proposed technical solution. The simulation of the block OR DNF-LUT in the system NI Multisim 10 by National Instruments Electronics Workbench Group presents Fig. 17-19.

Correctly, shaped value function and other combinations of k_1, k_2, k_3, k_4 , modeling confirms the efficiency of the unit disjunctions.

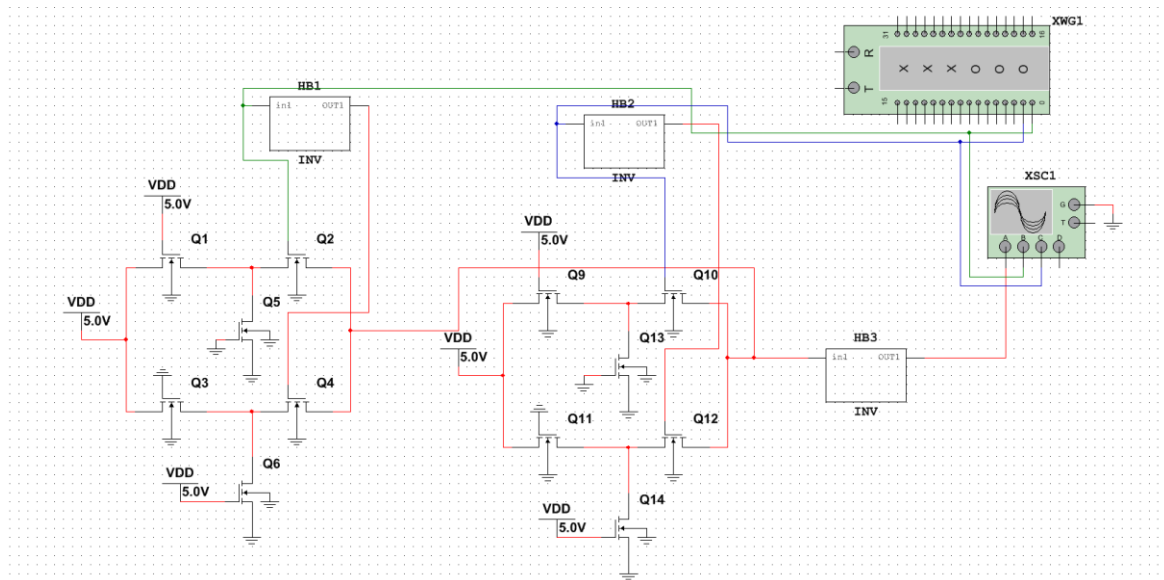


Figure 17. DNF – LUT - dynamic model

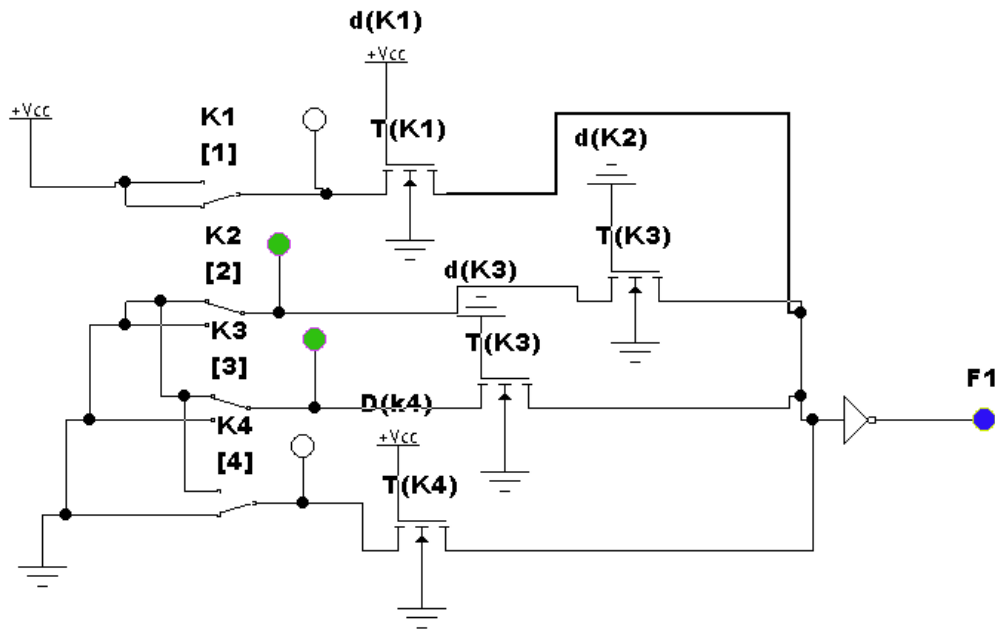


Figure 18. OR DNF-LUT Multisim model: Function $F_1 = 1$ as active k_1 and k_4 .

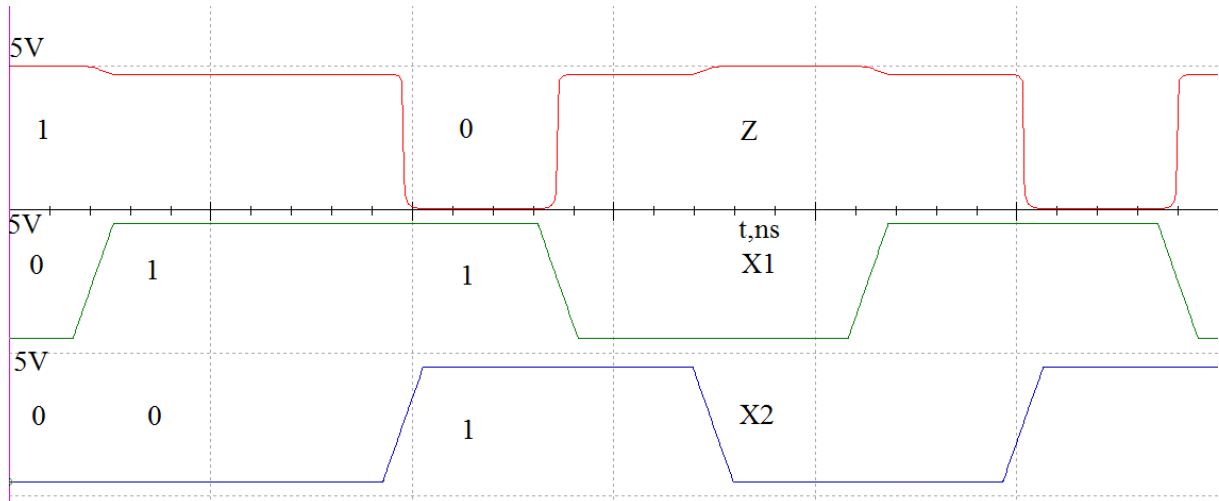


Figure 19. DNF – LUT - waveform

III. RESULTS OF CALCULATIONS

Let k - is the dimension of the main (basic) LUT ($k \in \{1, 2, 3, 4\}$). Let us estimate the complexity of the LUT without decomposition. "Ideal" complexity can be only up to $n=4$, not more than:

$$L_n = 2^n \cdot 8 + 2^{n+1} + 2n \quad (1)$$

With the decomposition n -tree for k LUT, we get "real" complexity:

$$L_{n,k} = 2^n \cdot 8 + (2^{k+1} + 2k) \cdot 2^{n-k} + (2^{2^{n-k} + 1} + 2^{n-k+1}) + 2n \quad (2)$$

Then we get the complexity of the two logical functions:

$$2 \cdot L_{n,k} = 2[2^n \cdot 8 + (2^{k+1} + 2k) \cdot 2^{n-k} + (2^{2^{n-k}+1} + 2^{n-k+1}) + 2n] \quad (3)$$

Using Double LUT, we get:

$$L_{d-n,k} = [2^n \cdot 8 + (2^{k+1} + 2k) \cdot 2^{n-k} + (2^{2^{n-k}+1} + 2^{n-k+1}) + 2n] + 2^{n+3} + 4 \quad (4)$$

In according with the expression (1-5) the complexity of the m known LUT described by:

$$L_{n,k} = m[2^n \cdot 8 + (2^{k+1} + 2k) \cdot 2^{n-k} + (2^{2^{n-k}+1} + 2^{n-k+1}) + 2n] \quad (5)$$

For DC LUT we get:

$$L_{dc-n,k} = [2^n \cdot 8 + 2 \cdot \{(2^{k+1} + 2k) \cdot 2^{n-k} + (2^{2^{n-k}+1} + 2^{n-k+1})\} + 2n] + \text{ceil}(\frac{n}{4}) \cdot (2k + 4m) \quad (6)$$

Considering expressions (4) and (6), taking into account the discrete parameters we get:

$$L_{dnf} = k(22n + 2) + m(k + 2) + 6m(k + \frac{n}{4} \cdot (2k + 4m)) + 2n \quad (7)$$

Where $k(22n + 2)$ - is a complexity (in the number of the transistors) of the AND blocks; $m(k + 2)$ - is the complexity of the OR blocks; $6mk$ - is the complexity of the OR blocks setting; $2n$ - is the complexity of the inverters for variable inputs; $\frac{n}{4} \cdot (2k + 4m)$ - is the complexity of the blocks in accordance with the restriction [11].

The comparison of the number of transistors to implement system of the logic functions in the known LUT and all proposed LUT at $m = 2$ are presented Fig. 20.

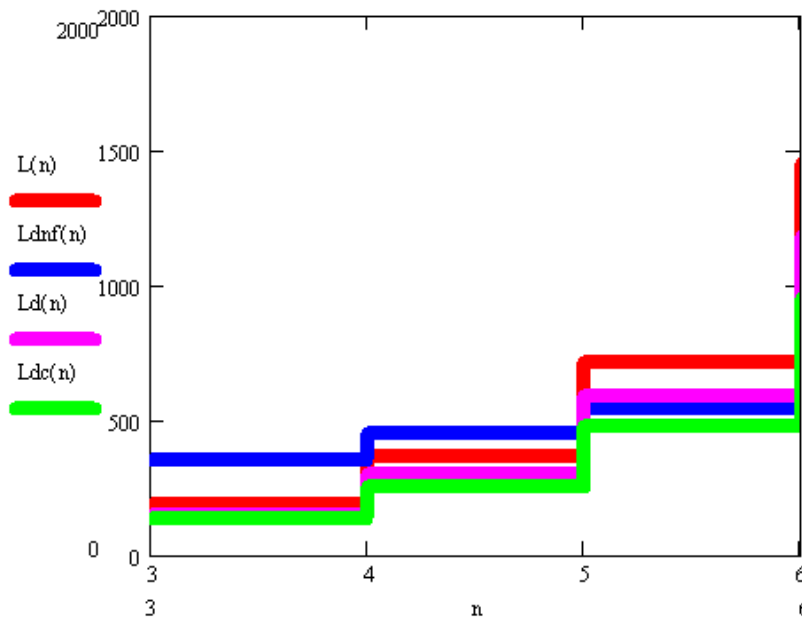


Figure 20. Comparison of the complexity of the m known LUT ($L(n)$) and proposed DLUT ($L_d(n)$), DCLUT ($L_{dc}(n)$), DNF-LUT ($L_{dnf}(n)$), $m = 2$

Thus a small number of functions inefficiently version DNF-LUT ($L_{dnf}(n)$). However, when calculating the 4 functions DNF-LUT ($L_{dnf}(n)$) ahead DCLUT ($L_{dc}(n)$) by 6 variables, Fig. 21.

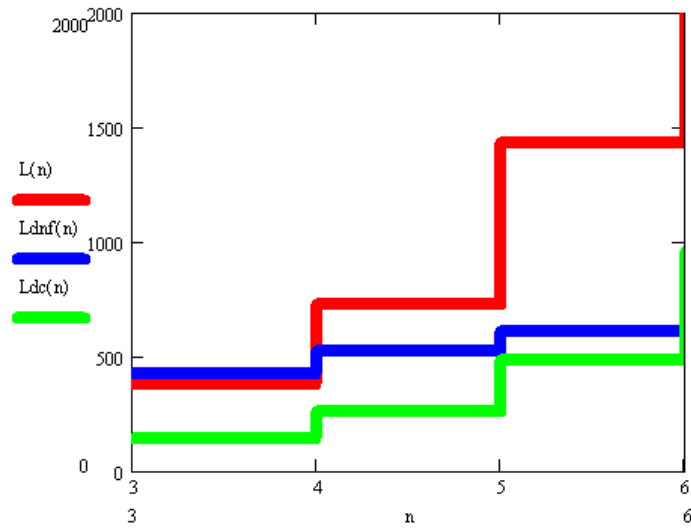


Figure 21. Comparison of the complexity of the m known LUT ($L(n)$) and proposed DLUT ($L_d(n)$), DCLUT ($L_{dc}(n)$), DNF-LUT ($L_{dnf}(n)$), $m=4$

DNF LUT is a competition about the number of variables is greater than 5, Fig. 22.

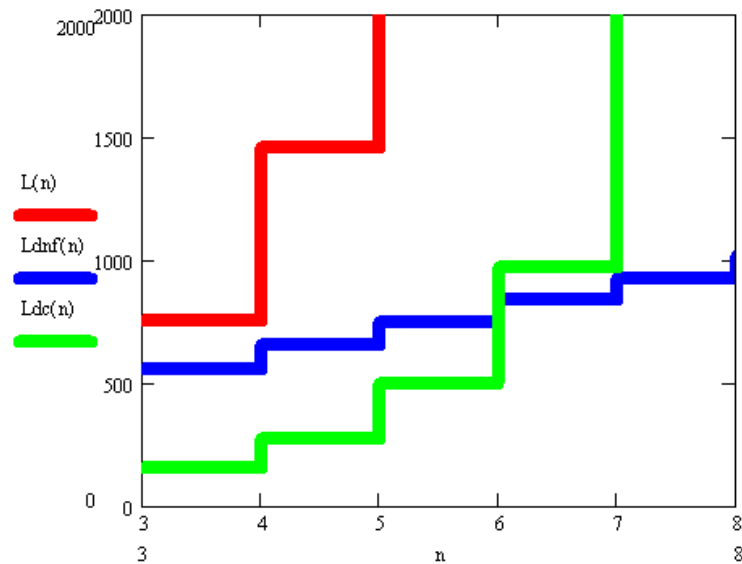


Figure 22. Comparison of the complexity of the m known LUT ($L(n)$) and proposed DLUT ($L_d(n)$), DCLUT ($L_{dc}(n)$), DNF-LUT ($L_{dnf}(n)$), $m=4; 3 \leq n \leq 8$

Comparison of the logical capacity (Number of transistors on a single function) of the m known LUT and proposed DCLUT, DNF-LUT presents Fig. 23, 24.

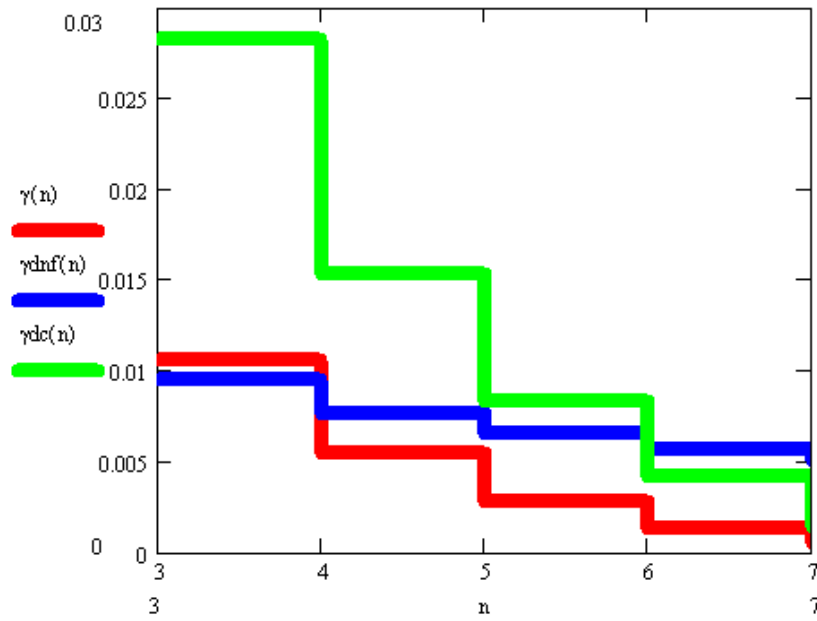


Figure 23. Comparison of the logical capacity of the m known LUT and proposed DCLUT, DNF-LUT, $m=4$

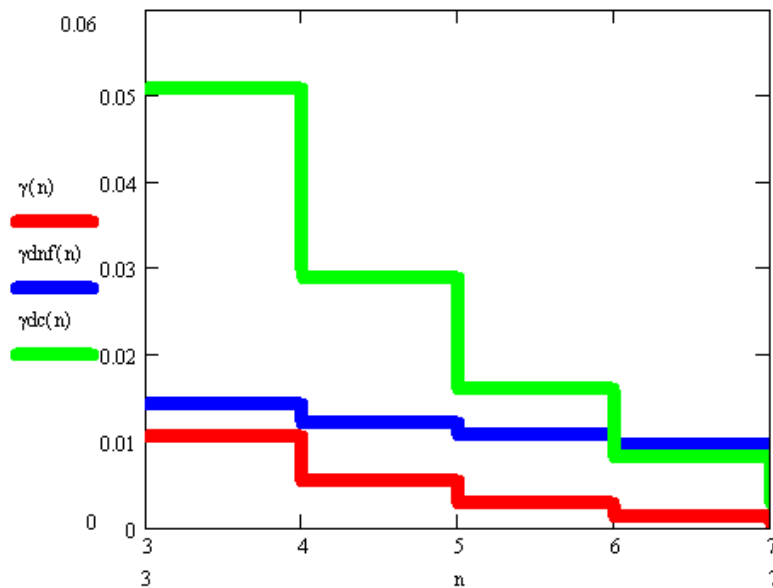


Figure 24. Comparison of the logical capacity of the m known LUT and proposed DCLUT, DNF-LUT, $m=8$

IV. CONCLUSION

We analyze the complexity of the proposed technical solutions and the results of functional simulation. Proposed advanced LUTs for the logic systems significantly reduces hardware costs (from 10% to 60% and more - Fig. 20-23) without essential reducing performance. The simulation of the advanced LUTs executed in the NI Multisim 10 by National Instruments Electronics Workbench Group and confirmed the efficiency of patentable technical solutions. If necessary, a small number of calculating functions appropriate to use DLUT. With an average number of logic functions, it is advisable to use the DCLUT. However, a large

number of variables DNF-LUT has the best characteristics as the complexity of the other options is growing exponentially. In the future, it is advisable to consider the integrated use of a variety of solutions in a single FPGA and perform appropriate optimizations. One of the promising areas of research may be the creation of adaptive LUT-based proposed advanced LUT.

REFERENCES

- [1] Tyurin S.F., Green Logic: Models, Methods, Algorithms. Green IT Engineering: Concepts, Models, Complex Systems Architectures, vol. 74, pp.69-86, 2017. DOI: 10.1007/978-3-319-44162-7
- [2] Tyurin S.F., Green Logic: Green LUT FPGA Concepts, Models and Evaluations. Green IT Engineering: Components, Networks and Systems Implementation, vol. 105, pp.241-261, 2017. DOI: 10.1007/978-3-319-55595-9
- [3] Drozd, A. et al.: The levels of target resources development in computer systems. In: Design & Test Symposium (EWDTS), 2014 East-West, pp. 1-5, IEEE, (2014).
- [4] Mehta, Nikil. An ultra-low-energy, variation-tolerant FPGA architecture using component-specific mapping. Dissertation (Ph.D.), California Institute of Technology. (2013) Available at: <http://thesis.library.caltech.edu/7226/1/Nikil-Mehta-2013.pdf>.
- [5] J. Cong, K. Minkovich. Optimality Study of Logic Synthesis for LUT-Based FPGAs. Available at: <http://cadlab.cs.ucla.edu/~kirill/tcad06.pdf>.
- [6] Robert J. Francis A Tutorial on Logic Synthesis for Lookup-Table Based FPGAs. Available at: <https://pdfs.semanticscholar.org/e879/aa350c38d381c41eeef0da95b739d2ede6c1.pdf>.
- [7] S. Brown. FPGA Architectura I Research: A Survey. Available at: <http://arantxa.ii.uam.es/~die/%5BLectura%20FPGA%20Architecture%5D%20FPGA%20architectural%20research.%20A%20survey.pdf>
- [8] S. Brown, J. Rose. Architecture of FPGAs and CPLDs: A Tutorial. Available at: http://malun1.mala.bc.ca:8080/~pwalsh/teaching/355/Toronto_tutorial.pdf
- [9] Tyurin S.F., Vikhorev R.V. Programmable logic device: patent RF №2573732; Published 27.01.2016, Bull. №3.
- [10] Tyurin S.F. Programmable logic device: patent RF №2544750; Published 20.03.2015, Bull. №8.
- [11] Conway, Lynn. Drafts of the Mead-Conway textbook, Introduction to VLSI Systems. University of Michigan. Retrieved 9 June 2015.

The system for control and accounting of working time

A. Plakhtii, V. Shkarupilo, S. Byelikov

Abstract—The comparative analysis of existing timesheet systems for creating a number of requirements to own solution has been conducted. Software requirements specification for working time control and accounting system (time tracking) has been created. The dependencies between the components of proposed solution have been analyzed. For this purpose, the Electron framework has been used. Basing on the results of the analysis conducted the system for working time control and accounting has been proposed. The proposed system is a cross-platform solution that allows to control and account the activity of employee during the working time.

Keywords—Timetracker, Timesheet Software, Working Time, Cross-platform.

I. INTRODUCTION

Working time is one of the most significant conditions of employment contract. Daily accounting positively affects the projects control aspects: it's much easier for project leaders to control the tasks accomplishments and the appropriate costs, to adjust the budget just in time (on the fly), with respect to current needs – not on the final stage of project-related work, when it's obviously that it's impossible to finish in time. Even the companies taking the fixed fee for the project work are beneficial here – they can easily analyze which categories of work and which clients are more preferable for them. The considered theme is especially topical for the people willing to work remotely and with respect to their own schedule – the freelancers [1].

The main task of working time control and accounting systems is to increase the business processes effectiveness, aimed at employees' work effectiveness increase – to give them the ability to establish the flexible schedules for themselves, with no time losses [2].

The working time accounting can be built on a daily, weekly or summarized basis. Nevertheless, in each of these cases the time spent on work per day is considered [3].

To estimate the peculiarities of employees' work effectiveness, to get the complete picture of how the working time is actually being spent, the software of the following classes is widely used: the time tracking software, the task management software and the PPS-systems (Personal Productivity Systems) which are also known as GTD-systems (Getting Things Done) [4].

There are the following reasons for time tracking systems usage: to involve the additional resources; to increase the accuracy of time utilization reports, to force the employees to be more responsible; to distinguish productive employees from the rest; to foster the new projects planning; to provide the control of remote freelance employees [5].

II. PROBLEM STATEMENT

The problem of employees' working time accounting becomes more and more topical each year, because to ground the illicit activities, carried out by the staffers during the working time, the continuous control has to be brought to the table [6]. Moreover, such control is intended to stipulate the employees to do their work more thoroughly and in more coordinated

A. Plakhtii Zaporizhzhya National Technical University, Zaporizhzhya, Ukraine (e-mail: n.plakhtii@gmail.com).
V. Shkarupilo Zaporizhzhya National Technical University, Zaporizhzhya, Ukraine (e-mail: shkarupilo.vadym@gmail.com).
S. Byelikov Zaporizhzhya National Technical University, Zaporizhzhya, Ukraine (e-mail: rector@zntu.edu.ua).

manner. Because of the fact that workers can operate with different operating systems, the problem of such universal cross-platform control and accounting system development arises.

The key aspect here is to diminish the costs intended to be associated with system tweaking with respect to the specific requirements of certain operating system. This peculiarity is a distinguishing feature of our work. To this end, the following tasks are resolved in our work: to conduct the comparative analysis of existing time tracking systems; formulate the requirements to own solution, with respect to the results of comparative analysis; to answer the question if our solution is actually a valid one – for this purpose the validation has to be conducted – by way of testing.

III. COMPARISON OF SIMILAR EXISTING SOFTWARE PRODUCTS

There are plenty of different time tracking systems. The typical one is a Timesheet Software – actually a table which should be manually filled on a daily basis – with a list of accomplished tasks, but only in electronic format [7]. Such software is typically installed locally, with a limited set of reports. It is applicable for small and medium companies in case the specificity of their activities doesn't cover numerous distractions, bounded with Internet and social networks in particular [8].

In a large companies there is a totally different picture – the vital need for Time Tracking (Recording) Software takes place. The appropriate systems should also be characterized with the following features: to provide the abilities to gather the information about the operations being performed by the employees at their workplaces; be integrated with modern project management systems [9]. An attempt to compare the feature sets of existing time tracking systems is given in Table 1.

TABLE I
THE COMPARISON OF MODERN TIME TRACKING SYSTEMS

Title Feature set	Harvest	Toggl	Rescue Time	Log My Hours	Prima ERP	Upwork
Manual / Automatic modes support	+	+	–	+	+	+
Differentiation between commercial and non-commercial projects	+	+	–	+	–	–
Automated reports generation	+	+	+	+	+	+
Creation of flexible invoices	+	–	–	+	+	+
An autonomous mode of timers functioning (automatic synchronization)	–	+	+	–	+	+
Functioning in a background with an active monitoring of application	–	–	+	–	–	–

Taking into consideration the content of Table 1, we assume that the most important functional properties of our system should be the following ones: the support of both – manual and automatic – modes of working time accounting; automatic reports generation with respect to defined time intervals; new projects have to be classified as commercial (profitable) and/or non-commercial (non-profitable) ones; in case of failures, the application has to be synchronized automatically.

Time tracking systems are accessible in various forms: local clients for Windows, Mac and Linux platforms; mobile clients; web services; extensions for web browsers; accounting systems for small business; free products and licensed software.

IV. SOFTWARE REQUIREMENTS SPECIFICATION

In accordance with software systems requirements specification drafting technique, proposed by IEEE (IEEE 830-1998 Standard), appropriately created specification has to provide customer with rigid and clear description of what is required to be done. For the executors such specification has to give the clear understanding of what the customer is expecting. For this purpose, the SRS (Software Requirements Specification) draft has to be elaborated. The structure and the content of concrete requirements specifications have to be defined. The additional supplementary documents should also be created, e.g. control letter for SRS quality checking and/or regulator's handbook [10].

The requirements specification for the system is briefly described below.

Introduction. The requirements specification describes functional and non-functional requirements to time tracking system. This document is devoted to be used by the company, responsible for system functioning correctness checking [11].

Our system is named as «TickTock». This system should allow the companies with about 100 employees, not taking into consideration the specificity of work, to account and control the activities of their staffers during the working time.

Time tracking system has to support the following functionality: the invoices accounting; to create the reports on weekly basis; to maintain the convenience of project management by way of task selection. To maintain these functionalities, the SMTP protocol (Simple Mail Transfer Protocol) has to be implemented. This protocol allows to transfer e-mails to postal server or from client computer, or between postal servers [12].

The further information is about the specifics – specific functional requirements and system limitations.

Detailed description. Our «TickTock» system is a plain time tracking system intended to be used by companies with about 100 employees, and their work has to be carried out on personal computers.

The system being developed reads and stores the metrics from workers' personal computers and builds the statistics from their activities, and also synthesizes the reports in form of electronic letter, then these reports are ought to be sent to the recipients with respect to the content of specified e-mails list.

User interface is represented with computer mouse controller, graphical interface with additional information and control elements: the information about working hours (spent / to be spent); tracking switch-on and switch-off buttons.

Software interface is represented with SMTP Gmail server – to send the messages (reports) to specified recipients.

Communication interfaces – with support of HTTP (Hyper Text Transfer Protocol) and SMTP protocols – to use the broadcast services.

Memory volume requirements – not less than 200 MB of free space on hard disk drive.

The usage requirements – the system has to be adopted for popular operating systems.

System functions. The system switches-on/off by a double clicking on utility icon. The first stage to get started is to log-in. After checking the login and password the following system's features will be available for the user: the time tracking start/stop button; the timer showing the time elapsed in hh:mm format; the ability to browse personal working statistics. If certain user is a super-user (with additional privileges), the following abilities are also allowed: to edit the information about users; to browse the statistics about users; to create the list of electronic addresses to send the statistical information about the employees' work; to edit the system settings (time zone settings, the limits of time intervals within which the tracking has to be conducted).

User characteristics. The client is a user (worker) which should have basic skills to work with basic windows applications.

The limitations. System can be utilized only by a registered user. The users can be registered only by a super-users – the administrators.

To make company be able to increase the number of registered employees the login and password have to be previously created for the first super-user. The system can be used on Windows and Linux operating systems. The Internet connection is a mandatory requirement.

Specific requirements. Only the registered user is allowed to be logged in.

Functional requirements to super-users: the ability to store time stamps when system is being switched on/off (after a specified period of time the snapshot of user screen is being taken and stored in a database); the availability of CRUD-functions set (create, read, update, delete) – to browse users' statistics, to shift time intervals; to modify the list of electronic addresses.

Functional requirements to users: to save the time when system is being switched on/off; when specified time is elapsed, the screen snapshot has to be taken and stored in database [13].

Quality requirements. From the safety viewpoint – the users authorization and authentication are mandatory [14].

V. SYSTEM IMPLEMENTATION AND TESTING

One of the most popular solutions for cross-platform applications development is Node.js framework, paired with Electron framework [15]. These frameworks can be easily paired up with any front-end JavaScript solution [16]. Our time tracking system has been implemented on this basis (fig. 1).

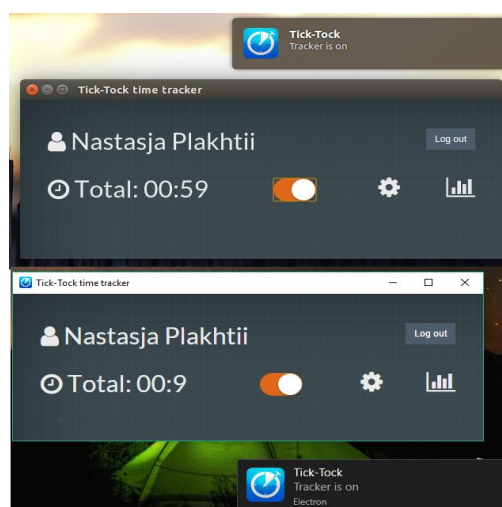


Fig. 1 Snapshot of the interface

In fig. 1 the snapshot of a system running on Ubuntu LTS 14.10 and Windows 10 PRO platforms is given to demonstrate the universality of our solution (to fulfill the universality requirement).

The intense time tracking system utilization in corporate environment is tightly bounded with appealing fault tolerance requirements [17]. The simultaneous connections to system by number of users can entail the faults in system functioning. To this end – to check the liveness properties of system – the Smart Bear LoadUI tool has been used [18]. This tool has been utilized to conduct the validation, and, thus, to prove the validity of system. The information about simultaneous connections and corresponding metrics are given in fig. 2.

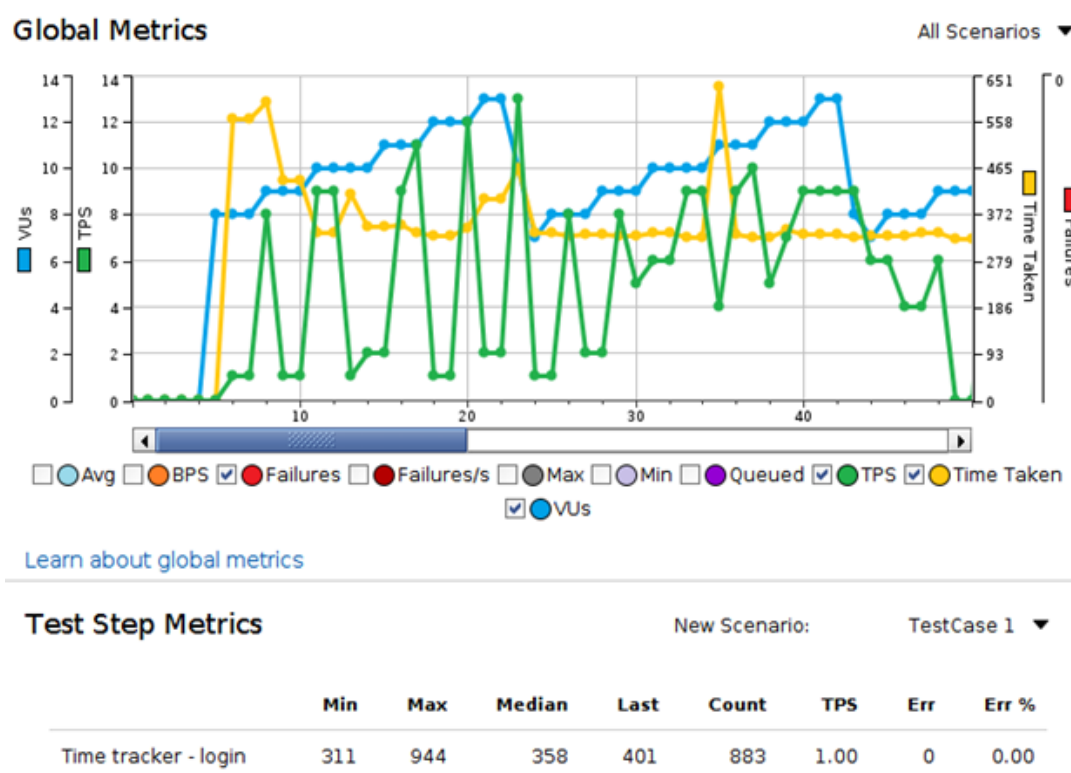


Fig. 2 Global metrics of testing process

In fig. 2 the following metrics can be seen: the maximum TPS (transactions per second) value – 13; Time Taken – the time (in milliseconds) taken for the requests (often comparable to response time values), the maximum value is about 620, the averages are about 300 ms; Failures – the number of failed requests – 0; VUs – the number of simultaneously running virtual users – 13 users. Taking into consideration the obtained results, it can be concluded that fault tolerance requirement is fulfilled and thus the proposed time tracking system can be characterized as valid one.

VI. CONCLUSION

Thus, the following tasks have been resolved:

1. The comparative analysis of existing control and accounting systems has been conducted. Basing on its results, a conclusion about system's most important functional properties has been formulated. The following properties have been distinguished as more important ones: the support of both – manual and automatic – modes of working time accounting; the ability to automatically generate reports for a given time intervals; the ability to differentiate between commercial and non-commercial projects; the automatic synchronization of application in case of failures.

2. Software Requirements Specification has been proposed. This allowed to define the requirements and limitations to be taken into consideration during the development process, e.g., universality, liveness.

3. The cross-platform solution has been developed. For this purpose the Electron framework has been used. The universality requirement fulfillment has been proved under Windows and Linux platforms.

4. The liveness requirement fulfillment has been proved by way of testing. It has been shown that proposed system is fault-tolerant and can be applicable to the scenarios with an intense workload – when multiple users (about 100) interact with system simultaneously.

Further work is aimed at integration of proposed solution with existing project management systems.

ACKNOWLEDGMENT

This publication is the result of the project implementation: TEMPUS CERES: Centers of Excellence for young REsearchers (Reg. no.544137-TEMPUS-1-2013-1-SK-TEMPUS-JPHES).

REFERENCES

- [1] Bundymuseum.org, “Willard Legrand Bundy Biography”, 2017. [Online]. Available: <http://bundymuseum.org/site3/about/the-history/willard-bundy-bio/>. [Accessed: 15-Dec-2017].
- [2] Skymark.com. “Frederick W. Taylor: Master of Scientific Management”, 2017 [Online]. Available: <http://www.skymark.com/resources/leaders/taylor.asp>. [Accessed: 15-Dec-2017].
- [3] Ko.com.ua, “Time Tracking Software - borcy s pozhiratelyami vremeni” (in Russian), 2013. [Online]. Available: http://ko.com.ua/time_tracking_software_borcy_s_pozhiratelyami_vremeni_81133. [Accessed: 15-Dec-2017].
- [4] Timetracker.biz, “Time Tracker” (in Russian), 2010. [Online]. Available: http://www.timetracker.biz/timetracker_function.html. [Accessed: 15-Dec-2017].
- [5] Blog-ru.primaerp.com, “Desyat' prichin dlya ispol'zovaniya tajm-trekerov” (in Russian), 2015. [Online]. Available: http://blog-ru.primaerp.com/2015/08/blog-post_26.html. [Accessed: 15-Dec-2017].
- [6] Coolwebmasters.com, “Luchshie MacOS programmy dlya otslezhivaniya rabochego vremeni dlya frilansera” (in Russian), 2012. [Online]. Available: <http://www.coolwebmasters.com/project-management/3313-time-tracking-mac-apps.html>. [Accessed: 15-Dec-2017].
- [7] Frey.su, “Trekeriy vremeni dlya programmista” (in Russian), 2015. [Online]. Available: <http://frey.su/programmer-time-tracker/>. [Accessed: 15-Dec-2017].
- [8] Onlineprojects.ru, “Time trackers” (in Russian), 2016. [Online]. Available: http://www.onlineprojects.ru/tags/tajm_trekery. [Accessed: 15-Dec-2017].
- [9] Lifehacker.ru, “5 trekerov vremeni, kotorye pomogut pobedit' prokrastinaciyu” (in Russian), 2016. [Online]. Available: <https://lifehacker.ru/2016/11/17/trekery-vremeni>. [Accessed: 15-Dec-2017].
- [10] IEEE Std 830-1998. IEEE Recommended Practice for Software Requirements Specifications. The Institute of Electrical and Electronics Engineers, Inc., 1998. ISBN 0-7381-0332-2.
- [11] Dpgrup.ru, “Trebovaniya k programnym produktam” (in Russian), 2017. [Online]. Available: <http://www.dpgrup.ru/software-requirements.htm>. [Accessed: 15-Dec-2017].
- [12] M. Miller, *Sams Teach Yourself...in 10 Minutes*. US: Pearson Education, 2014.
- [13] K. Adams, *Non-functional Requirements in Systems Analysis and Design*. US: Springer International Publishing, 2015.
- [14] S. Han, G. Skinner, V. Potdar, and E Chang, “A framework of authentication and authorization for e-health services,” in *Proc. 3rd ACM workshop on Secure web services*, Alexandria, Virginia, USA, 2006, pp. 105–106.
- [15] Electronjs.org, “Build cross platform desktop apps with JavaScript, HTML, and CSS”, 2017. [Online]. Available: <https://electronjs.org>. [Accessed: 15-Dec-2017].
- [16] T. Brown, *Web Development with Node and Express*. US: O'Reilly Media, 2014.
- [17] V.V. Gribova, and A.S. Kleshchev, “Metody i sredstva razrabotki pol'zovatel'skogo interfejsa: sovremennoe sostoyanie,” *Software & Systems*, vol. 1, pp. 1–6, 2001.
- [18] Smartbear.com, “Behind Every Piece of Quality Software is a Great Software Team”, 2017. [Online]. Available: <https://smartbear.com>. [Accessed: 15-Dec-2017].

Comparative Statistical Analysis of Pseudorandom Binary Sequences by 53-bit Internal State Size Generator

Tatiana Menshikh, Ilya Piatrenka, Vitaliy Khazan, Stanislav Derechennik

Abstract—In this paper we present a comparative statistical analysis of the subset of pseudorandom binary sequence generators. The Micali-Schnorr, G-SHA-1 generators and unique 53-bit internal state size generator are compared using NIST Statistical Test Suite. For practical purposes statistical tests allow extracting data on the generator producing a truly random sequence. 15 tests for sequences were carried out and analyzed.

Keywords— Random number generation, testing, statistical analysis.

I. INTRODUCTION

Pseudorandom binary sequence (PRBS) generators may be used in many applications and they are central to cryptographic protocols and algorithms development. PRBS-generators testing is important both theoretically and practically. To analyze if the sequence is random, it is necessary to determine the quality of generators. Randomness is a probabilistic observation, so the sequence is evaluated on the basis of the probability theory. A variety of statistical tests can be applied to pseudorandom sequences. The best known are the NIST suites of statistical tests (NIST STS) [1], Donald Knuth's test [2], the DIEHARD [3] and the Crypt-XS [4].

The NIST suite of statistical tests in the package include the following 15 tests: frequency, block frequency, cumulative sums (2 subtests), runs, long runs, Marsaglia's rank, spectral (based on the Discrete Fourier Transform), non-overlapping template matchings (148 subtests), overlapping template matchings, Maurer's universal statistical, approximate entropy (based on the work of Pincus, Singer and Kalman), random excursions (due to Baron and Rukhin, 8 subtests), random excursions variant (18 subtests), linear complexity, and serial (2 subtests) [1]. The basis ideas of these tests are those of Donald Knuth from his book [2], and also the DIEHARD tests developed by George Marsaglia [3]. However, all NIST tests are in the uniformed computational technique consisting of four steps. Firstly, a specific null hypothesis is formulated as the main hypothesis. Secondly, P-value statistics are calculated for the fixed block length of pseudorandom binary sequences. Thirdly, the uniformity of the resulting P-value distribution in the range $[0,1]$ is verified. Fourthly, the relative frequencies of the P-value exceeding significance level (α) with the subsequent verification of this value within the confidence interval $[\beta_{min}, \beta_{max}]$ depending on α . If P-value is within this interval, then null hypothesis is not rejected, and it means that the sequence passed the test successfully.

To compute probabilistic P-values, there are used some special functions in NIST STS: Standard normal (Cumulative Probability Distribution),

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z \exp(-u^2/2) du \quad (1)$$

T. J. Menshikh Brest State Technical University, Brest, Belarus (e-mail: menshikh93@bk.ru).

I. A. Piatrenka Warsaw University of Life Sciences, Warsaw, Poland (e-mail: petrenko.ilia@gmail.com).

V. L. Khazan, Omsk State Technical University, Omsk, Russia (e-mail: vlhazan@yandex.ru).

S. S. Derechennik Brest State Technical University, Brest, Belarus (e-mail: cm@bstu.by).

Complementary error function,

$$\operatorname{erfc}(z) = \frac{2}{\sqrt{\pi}} \int_z^{\infty} \exp(-u^2) du \quad (2)$$

Incomplete gamma function,

$$Q(a, x) = 1 - P(a, x) \equiv \frac{\Gamma(a, x)}{\Gamma(a)} \equiv \frac{1}{\Gamma(a)} \int_x^{\infty} \exp(-t) \cdot t^{a-1} dt \quad (3)$$

where $Q(a, 0) = 1$, $Q(a, \infty) = 0$, and a is a variable parameter computed for each test individually [1].

II. PRELIMINARIES

Three generators were chosen for testing. The first one is a linear congruential generator previously developed and described by authors – the 53-bit internal state size generator G53 [5], [6]. Second and third generators are cryptographically secure G-SHA-1 and Micali-Schnorr (MS) ones [1]. 53-bit generator produces pseudorandom number sequence in the range $[0, 1]$, as illustrated in Fig. 1, and each pseudorandom number x_i for this number sequence is changed to binary b_i

$$b_i = \operatorname{fix}(2 \cdot x_i) \quad (4)$$

where fix is a function which rounds $2 \cdot x_i$ value to the nearest integer toward zero.

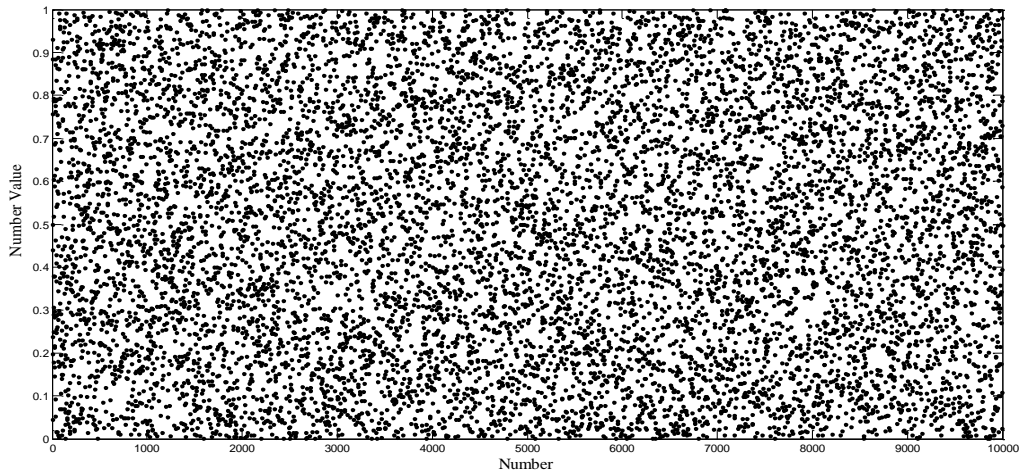


Fig. 1 The plot of pseudo-random numbers for 53-bit Generator

These generators have long M-sequences. For example, G53 generator have the period $M = 2^{53} - 1$. For each generator testing, the sequence length is $N = 2^{20}$ bit, the sample size is $n = 256$, the significance level $\alpha = 0.01$.

The parameters for NIST tests are shown in Table I.

TABLE I
NIST TEST PARAMETERS

Statistical Test	Block length, bit
Block Frequency	16384
Non-overlapping Template	9
Overlapping Template	9
Approximate Entropy	10
Serial	16
Linear Complexity	500

Two basic checking methods of P-value distribution uniformity are used. First one is estimation of expectation (mean), variance estimate computing, and their comparison with theoretical: mean $m = 0.5$, variance $v = 0.0833$. Second one is chi-square using P-value histogram analysis.

III. ANALYSIS AND COMPARISON

The comparative analysis results for tested generators as percent deviations of mean and variance from theoretical results are shown in Table II.

TABLE II
MEAN AND VARIANCE PERCENT DEVIATION

Statistical Test	MS		G-SHA-1		G53	
	$m(\%)$	$v(\%)$	$m(\%)$	$v(\%)$	$m(\%)$	$v(\%)$
Frequency	4.34	4.68	4.50	12.00	8.66	6.48
Block frequency	4.70	4.80	1.48	4.80	3.44	0.72
Cumulative sums	4.28	1.32	0.54	5.40	8.78	0.12
Runs	5.80	7.20	2.06	8.76	6.76	2.52
Long runs	5.22	1.20	1.44	1.56	3.74	6.96
Rank	3.36	1.68	3.6	0.84	0.66	2.52
Spectral	1.06	10.68	0.56	1.20	2.50	6.72
Non-overlapping template	0.10	0.84	0.04	0.36	0.22	0.48
Overlapping template	3.30	16.21	3.42	1.08	2.34	3.84
Universal statistical	0.30	0.84	4.34	0.84	8.84	2.28
Approximate entropy	0.28	7.68	0.02	0.48	12.34	22.2
Random excursions	0.04	0.84	1.44	0.24	1.78	0.00
Random excursions variant	1.22	2.88	0.72	4.56	0.36	2.16
Serial	3.70	13.69	4.10	0.24	9.58	2.28
Linear complexity	6.54	8.52	1.94	7.32	7.92	7.56

The minimum location parameter (expectation) deviations are observed for G-SHA-1 generator – 8 from 15 tests, for Micali-Schnorr generator – 4 from 15 tests, and for G53 generator – 3 from 15 tests. The minimum scale parameter (variance) deviations are noticed in 7 from 15 tests for G-SHA-1 generator, 5 from 15 tests for G53 generator, and 3 from 15 tests for Micali-Schnorr generator.

To show the distribution uniformity, histograms are made for one of the tests, non-overlapping template matching test, as illustrated in Fig. 2.

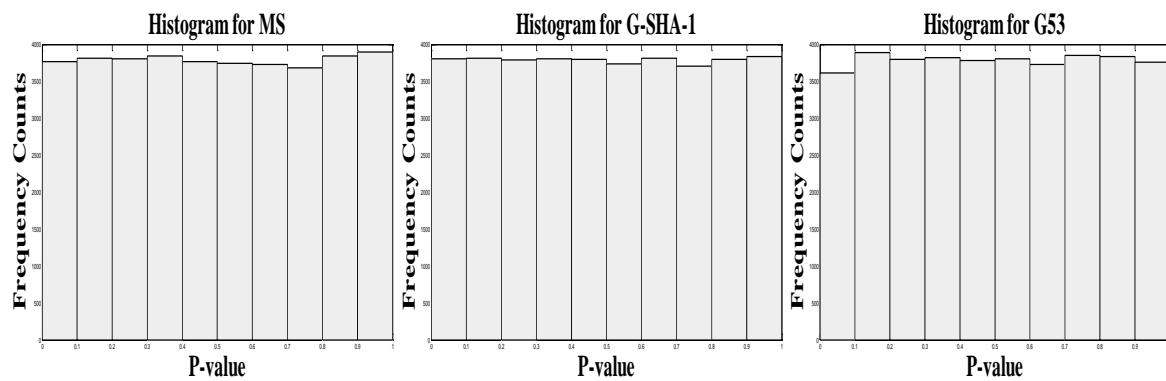


Fig. 2 Non-overlapping template matching test P-value histograms for Generators

To estimate the P-values on the significance level $\alpha = 0.01$ (about 1% of the sequences are expected to fail), the confidence interval is defined as $\beta \pm 3\sqrt{\beta(1-\beta)/n}$, where $\beta = 1 - \alpha$, and n is the sample size. For our sample size $n = 256$, the confidence interval is 0.99 ± 0.018656 . For each statistical test, the proportion of sequences that pass is computed. This is illustrated on Figs. 3–5 for all three generators (with straight horizontal line making the confidence interval lower boundary).

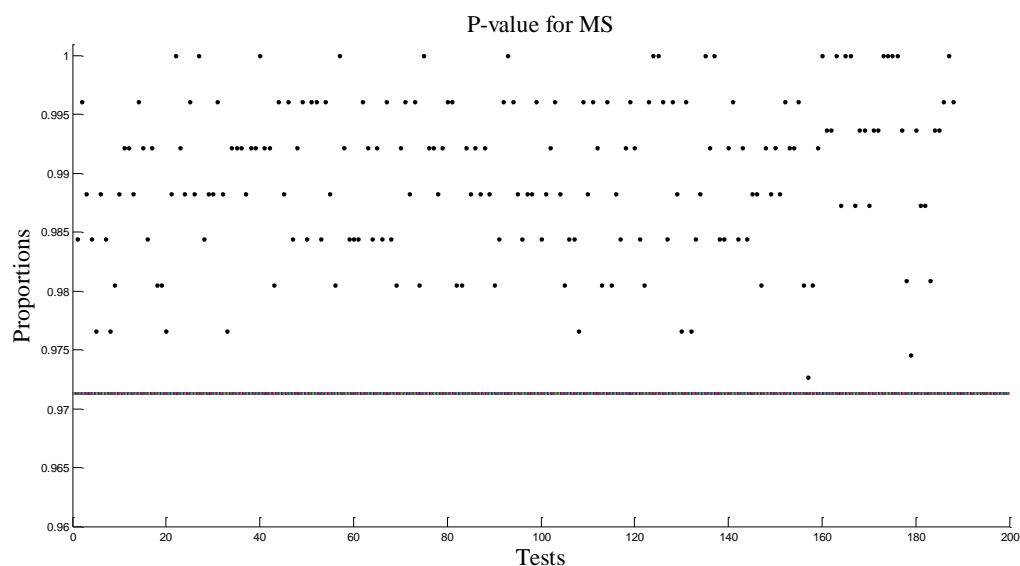


Fig. 3 P-value plot for Micali-Schnorr Generator

As can be seen in Fig. 3, the computed P-values for all tests are within the confidence interval. For G-SHA-1 generator, the significant deviation for the 44-th subtest of non-overlapping template matching test was revealed. For G53 generator, less significant deviations for the 23-th and 93-th subtests also for non-overlapping template matching test for the calculated confidence interval were revealed.

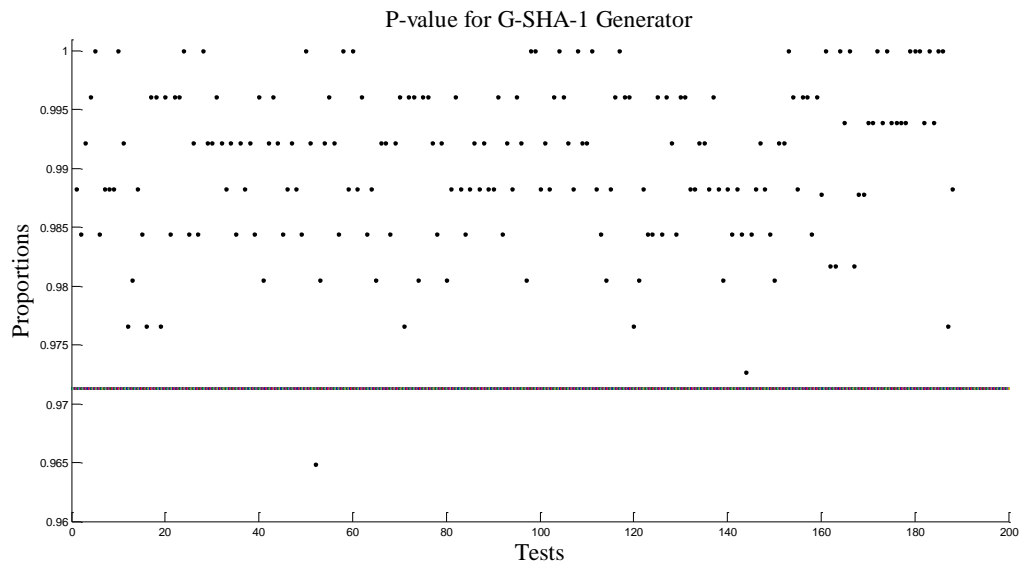


Fig. 4 P-value plot for G-SHA-1 Generator

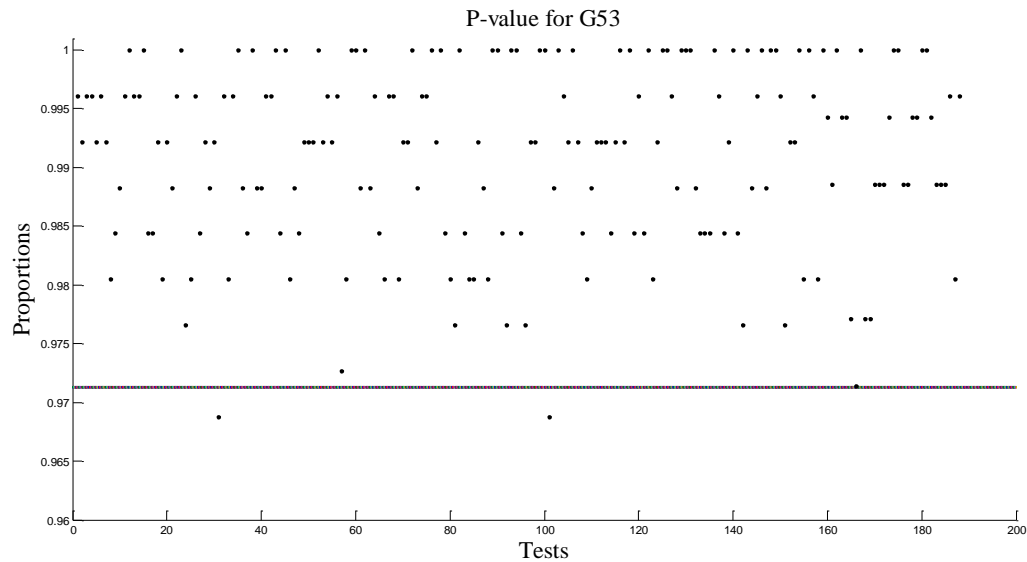


Fig. 5 P-value plot for G53 Generator

Therefore, the operating results for all three generators can be considered as satisfactory.

IV. CONCLUSION

The presented results of the pseudorandom binary sequence generators comparative analysis are showing that 53-bit generator can compete with the higher class generators, at least with G-SHA-1, for requirements comparable with ones researched (sequences length $N = 2^{20}$ bit and the sample size $n = 256$). In the same time G-53 is obviously faster and simpler in implementation due to its congruency, and this fact makes it a viable choice for a large subset of practical tasks.

REFERENCES

- [1] A. Rukhin, J. Soto, et al, "Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications," NIST, US, 2010.

- [2] D. E. Knuth, "The Art of Computer Programming, Seminumerical Algorithms," 3rd ed. vol. 2, Addison Wesley, Reading, Massachusetts, 1998.
- [3] G. Marsaglia, "A current view of random number generators," Proc. Comput. Sci. Statistics: Sixteenth Symp. Interface, Mar., 1985.
- [4] W. Caelli, "Crypt x package documentation", Tech. Rep. Information Security Research, 1992.
- [5] V. L. Khazan, "Mathematical Models of Discrete Decameter-Wave Communication Channels," Publishing House of Omsk State Technical University [in Russian], Omsk, 1998.
- [6] T. J. Menshikh, "Pseudorandom number generators for cryptographic secure of communication channels," in Proc. 9th Young Scientists and Students Conf. [in Russian], Brest, 2015, pp. 31–34.

Fostering Competences Development in Belarussian Higher Education

U. A. Beizerau

Abstract – The research deals with processes of modernization of Belarussian higher education and use of the competences' approach in reforming of educational system. Important issues of integration of Belarussian educational system into the world educational space is also discussed. Concepts of modernization of national system of education are and the process of change of an educational paradigm which is understood as initial conceptual model of statement of problems and their solvation by means of certain methods are highlighted in the article. Erasmus+ project FOSTERC (Fostering competences' development in Belarussian higher education) is described as an example of an instrument for modernization and reforming educational systems.

This publication is done in the frameworks of the ERASMUS+ project "Fostering Competencies Development in Belarussian Higher Education" (FOSTERC) Project Ref. N° 574087-EPP-1-2016-1-ES-EPPKA2-CBHE-SP which is supported by the European Commission. (Dr. Adela Garcia-Aracil, Polytechnic University of Valencia is the coordinator and grantholder). The application materials and preliminary results are used in the article.

Keywords – competences; integration; the system of education; educational space; globalization; processes of integration; modernization of education.

I. INTRODUCTION

Higher education is characterized today by the processes of globalization, humanization and fundamentalization directed to the formation of the person as a cultural and historical subject, capable to productive conversion and implementation of potential of culture in specific socially and professionally significant qualities of personality (competences). It is connected with a new role of education in the modern world, which draws the important task – an advancing designing of the new social relations in the context of actively developing global economy, the global information environment, cross-cultural dialogue, internationalization and standardization of various spheres of activity of people of different countries of the world. The changes in nature of education, its orientation, purposes and content happening in the World and drive it to free development, creative initiative and independence of students, competitiveness and mobility of specialists.

II. MODERNIZATION OF HIGHER EDUCATION

Concepts of modernization of education are united worldwide by the process of change of an educational paradigm which is understood as initial conceptual model of statement of problems and their solvation by means of certain methods. The new educational paradigm is oriented to forming of core competencies of the specialist as a result of formation in total of its motivational, valuable and cognitive components [2].

In the conditions of the accelerating changes in social, political, legal, economic and technological spheres of life higher education institution is urged not only to give professional knowledge, skills, but also to create a number of professionally significant and socially necessary personal qualities (competences) of the specialist. Among the main competences are the following: culture of system professional thinking, communicative culture, ability to work in team, tolerance, aspiration to self-education and self-development, responsibility, organizing and leadership skills, resistance to fixed changes (social and the environmental), flexibility and creativity of thinking, optimum style of professional behavior, ability to represent personal and professional qualities, culture of a healthy lifestyle [4].

At modernization of higher education, European countries consider the principles and approaches which are contained in the international documents on education, in particular in the Bologna declaration on June 19, 1999 which so far was signed by 40 states. According to this document by 2010 all European countries had to create uniform system of higher education with the uniform

educational standard and the recognized diplomas of university graduates in all State Parties of Bologna Process [1].

Basic points of the Bologna declaration are: acceptance of system of easily understood and comparable degrees (graduates can use effectively the qualifications in all countries of the European higher education); establishment of the system of credit units in training process); social and professional mobility of students and teachers; providing standards of high quality and comparability of qualifications in all European countries; life-long training; involvement of higher education institutions and students as its competent, active and constructive partners in forming and development of the European higher education space.

These approaches are correlated to the global strategies of development of education developed by UNESCO in the 21st century according to which the most important functions of educational institutions are the following: "to learn to live together", "to learn to learn", "to learn to do", "to learn to live". Basing on the principles, listed above, the Council of Europe determined five core competencies which must be gained by graduates of universities: social and political; cross-cultural; communicative (oral speech and written communication native and foreign languages; connected with emergence of information society; connected with a capability to study throughout all life in the context of the professional activity [5].

All the competences, listed above, form social and professional competence of the specialist. The analysis of essence of this concept shows that the focus in it is shifted from the conceptual field "I know that" to the field "I know how". In the Concept of reforming of Belarussian education it is also specified the period that we shall create complete system of encyclopedic knowledge, abilities, skills, and also experience of independent activities and the personal liability of students, i.e. the core competencies determining modern quality of education.

III. COMPETENCE-BASED APPROACH IN EDUCATION

The education (competence-based education) oriented to competences began to be created in the 70th of the 20th century in the USA. Foreign and domestic researchers of competences (J. Raven, C. Scala, R. Whyte, V. I. Baydenko, G.E. Belitskaya, E. F. Zeer, I. A. Zimnyaya, N. V. Kuzmina, A. K. Markova, L. A. Petrovskaya, Yu. G. Tatur, G. Halazh, N. Chomsky, V. Hutmakher, A. V. Hutorskoy) connect this concept, first of all, with categories "readiness", "capability", and also with such psychological qualities as responsibility and confidence. In the list Ravenna is 37 types of competences which many names include the words "readiness" and "capability", for example, "readiness and a capability to study independently", "readiness to use the new ideas and innovations", "a capability to make decisions", "a capability to joint operation for the sake of goal achievement" [4].

On the basis of the analysis of results of many researches of competence-based approach in education it was theoretically proved and allocated three groups of core competencies, their necessary nomenclature and types of competences from into each group. The first group is constituted belonging to the person as the identity of competence: health-saving; valuable and semantic orientation in the world; civic consciousness; self-improvement, self-regulation, self-development, personal and subject reflection. The second group is formed belonging to social interaction of the person and social environment of competence: social interaction with society, labor collective, family, friends, cooperation, tolerance, respect and acceptance of others, social mobility; in oral and written communication. The third group includes the competences relating to activities of the person: cognitive activity, including statement and the problem resolution; educational, game and labor activity; in the sphere of information technologies, including computer literacy and of Internet skills [6].

The analysis of tendencies of development of higher education in the world demonstrates that increase in its quality and efficiency depends not so much on curricula and the number of classroom hours, but on the identity of the teacher capable to realize the installations, knowledge, skills in the creative potential of students. Training of such teachers who are capable to build cross-disciplinary

curricula, enter productive communication with other teachers, administration and students, use various styles of pedagogical activities, constantly improve the skills is an extremely topical issue today. Development of productive educational activities, autonomy of the personality and creativity of students, naturally, demands that the teacher himself had pronounced pedagogical identity, creativity, ability to flexible professional behavior, creative search and professional self-improvement [8].

Necessary condition and means of self-improvement of the identity of the teacher is his constant reflection of the professional activity. Development of a professional reflection, adequate self-assessment becomes today one of priority problems of vocational training and professional development of teachers without which solution or any essential modernization of higher education is hardly possible. As the instrument of measurement of level of formation of professionally significant qualities of the identity of the teacher, key competences, development of his reflexive and estimated consciousness it is necessary to use widely the diagnostic techniques and technologies developed in pedagogical medical ethics and psychology [9].

Today, when Belarussian education is integrated into European and World educational space, and joined Bologna process, it is extremely important to preserve and even improve the quality of higher education and make Belarussian higher educational institution competitive in the market.

Erasmus+ project FOSTERC (Fostering competences' development in Belarussian higher education), which started in 2016 and united 11 Belarussian HEIs became one of the important instruments of integration and competencies' formation.

FOSTERC is directly focused on the National (and Regional) priorities fixed by Government of Belarus, regarding the modernization of governance, management and functioning of Belarusian higher education system. Immediately upon attaining independence, Belarus has made a breakthrough towards the formation of a socially oriented market economy. Belarus has adopted the concept of sustainable development as the ideology of its social, economic and cultural development. The changes taking place in the economic and sociocultural life of the country and the realities of the new historical phase of development are having an impact on the development of Belarusian education system. The creation of the legal and economic mechanisms to ensure the functioning and development of the education system has made it possible to give effect to the constitutional right of citizens of Belarus to receive high-quality and accessible education. The fact that Belarus has achieved some of the Millennium Development Goals ahead of schedule, the high level of human development and the ensuring of political, socio-economic and inter-ethnic stability within the country are evidence of the results achieved (UN, Economic and Social Council, 2011).

In this context, one of the objectives, fixed for the National Strategy for Sustainable Socio-economic Development of Belarus up to the year 2020, is to bring the national higher education system up to a level that matches world standards, while retaining its national and cultural identity. In this sense, particular attention is paying to increasing the quality of higher education and improving the structure for training specialists and bringing it as close as possible to the requirements of the labour market. The development of competent research and teaching faculty matching international requirements are pushing through increasing academic mobility, skills development and teacher training, including internships abroad that are providing access to new knowledge and technology. Moreover, in line with world trends, the quality of higher education and the recognition of national institutions of higher education are improving with the completion of the process of obtaining ISO 9001 certification for the educational and research activities of Belarusian universities and by implementing elements of the Bologna process in the national system of higher education.

IV. FOSTERC PROJECT AND ITS ROLE FOR BELARUSSIAN EDUCATION

However, although higher education in Belarus has a good level, the recent integration of Belarus in the Bologna process requires some harmonisation with other higher education system in Europe. The changes in the higher education system should be focused on the emphasis of the use of innovative principles and approaches to teaching and learning, the strengthening of resources

allocation and methodological support of education, the upgrade and improvement of the technical base, the use of information technologies, among other issues. In this framework, FOSTERC project supports the process of convergence focusing, in this case, on learning outcomes, one of the most important aspects to develop in order to increase the quality of the system, the management of teaching and learning processes and eventually the connection with the social and economic demands.

The added value of FOSTERC is that it provides to Belarusian universities critical tools for reaching Belarusian priorities, such as: training university staff how to define and validate learning outcomes; promoting cultural changes among academics and stakeholders; providing some economic incentives to start up the first initiatives for implementation and evaluation of student-centred teaching models based on competencies development. Along with, FOSTERC will make possible the modernization of Belarusian public universities (and hopefully the modernization of the rest of HEIs) through one of the most powerful tools for changing universities: the shift towards using traditional input/content to output/outcomes approaches in Belarusian HEIs (i.e. opening up ivory towers to social demands).

FOSTERC project is being viewed as producing innovations in a public good sense. FOSTERC not just develop and apply a methodology for own purposes. The methods will be documented and codified to facilitate their potential use by others in the future. These include the overall conceptual, empirical and evaluation framework, the specific quantitative (graduates and teachers questionnaires) and qualitative (employers interviews) analysis, the training material of workshops and EU visits, the design of a student-centred teaching model based on competencies development and its implementation through a pilot activity in Belarusian participating universities, the definition of some Belarusian HE learning outcomes and the development of an integrated «Management Information System», and the learning process for policy makers through different events proposed in the project and the Policy Recommendation Guide. Moreover, members of the consortium are actively involved in several international networks that will enable efficient and thorough exchange of the FOSTERC findings with the results of other relevant projects.

The main aim of FOSTERC project will be to strengthen the use of innovative principles and approaches to teaching and learning in the Belarusian Higher Education Institutions (HEIs) for the improvement of graduates' learning outcomes. FOSTERC will examine the curricular development focusing on graduates' competencies (knowledge, skills and attitudes). It will foster competencies development for matching the social and economic demands and improve the overall quality of HE. Implementing these reforms, it will support the modernization of governance, management and functioning of Belarusian HEIs, and the upgrade and improvement of teaching and learning methodologies. It will provide trainings, methodological tools and indicators for conceptualization of learning outcomes, and integrate good practices and new methods into daily activities. It will also provide the necessary information and tools to facilitate the implementation of internal and external quality assurance assessments, supporting, at the end, the modernisation of Belarusian participating universities and hopefully the modernisation of the rest of Belarusian HEIs.

Specifically, FOSTERC project supports the modernisation of governance, management and functioning of Belarusian HEIs through the following objectives: providing indicators for conceptualizing Belarusian HE learning outcomes in terms of competencies development; upgrading and improving innovative teaching methods and the development of pedagogical creativity; strengthening student's assessment and education management information systems for the improvement of the quality of Belarusian HEIs and matching the social and economic demands; equipment of Belarusian participating HEIs with the necessary information tools to facilitate the implementation of internal and external quality assurance assessments; modernizing systems and practices of data collection and analysis in the Belarusian HE system through the development of an integrated Management Information System improving HEIs management in teaching and learning activities.

All these objectives will positively change the situation set out previously and will improve the overall quality of Belarusian universities.

FOSTERC project is quite intensive in human resources. Belarusian participating universities are responsible for conducting HE graduates and teachers surveys, and interviewing employers to know the current situation of teaching approaches, learning outcomes and labour market demands in Belarus. In addition, there are many training seminars in Belarus and EU visits and they require the cooperation of many people. Moreover, the definition of the student-centred teaching model based on competencies development, their establishment, implementation and validation in the Belarusian participating universities (these activities correspond to implies an inter- and intra-institutional cooperation among education and training practitioners, institutional managers and policy-makers on the relevance of using learning outcomes for defining standards, curriculum development and assessment/validation. Furthermore, activities of the project require involving national and international external experts giving support to Belarusian teams and improving project quality results.

To maximize cost effectiveness and to ensure that the project is manageable, numerous measures have been taken: 1. Staff costs are estimated according to number and volume of activities.

2. Type of task is used as indicator of the daily salary rate. 3. Training for Belarusian staff will be held mostly in Belarus. 4. When traveling, the duration of visits is planned in regard to activities. Travel costs have been reduced as much as possible. 5. Reasonable prices are estimated for equipment. 6. To reduce costs for printing and publishing, apart from institutional reports and a publication on major results, dissemination will be mainly done via project website.

7. Translation services from Russian and Belarusian to English have also be considered for reports, seminars, training EU visits and Final Conference. 8. A financial expert in EU projects is in charge of the management of the project. Financial administration for the project will be centralised in order to reduce costs both for book-keeping staff and international external experts collaborations, and for financial transactions. Detailed record will be kept for financial reporting and audits.

The expected impact of the project can be divided into three levels:

LOCALLY: FOSTERC is aimed at promoting the modernisation and internalisation of Belarusian HEIs. In concrete terms, this can be achieved by providing: a comprehensive picture of the current situation of teaching and learning approaches of the Belarusian HEIs and the process of moving towards the implementation of the Bologna process; a better understanding of the curricular development focusing on learning outcomes; a better matching of the Belarusian HE with the social and economic demands; an increasing use of learning outcomes when describing and defining qualifications, and curricula; quality assurance assessments tools for Belarusian HEIs enhancing its relevance for the labour market and society.

All information and data collected throughout the project helps for better understanding and elaboration of that.

NATIONALLY: FOSTERC advances our knowledge in how fostering competencies development in Belarus. Our proposed mixed methods approach reflects the objectives of our study and of the call, consisting of qualitative and quantitative analysis, questionnaire surveys, training workshops and EU visits, defining student-centred teaching models based on competencies, implementing full pedagogical models in some study programmes of Belarusian HEIs, updating and integrating teaching methodologies, and improving quality assessment of Belarusian HE throughout the recognition and validation of competencies. These approaches are be conducted for each Belarusian participating university. As FOSTERC covers 8 Belarusian HEIs from different geographical areas with a broad view of cultural diversity and representing well the Belarusian context, each analysis of project activities will produce important results in its own right, but will also provide input to both participating and non-participating Belarusian HEIs. Therefore, the impact of the project is expected to be the desired one increasing synergies and complementarities among all Belarusian HEIs.

EUROPEAN LEVEL: FOSTERC will deliver large scale evidence of the provision and assessment of basic and transversal skills, highlighting those related to entrepreneurship, social civic, intercultural competencies, critical thinking, etc. Conducting HE graduates and teachers surveys, and employers interviews, their analysis and the integration of evidence into policy making process and

the use of the learning process approach for the policy makers and practitioners, will help for a better understanding of which are the key competencies linked to professional profiles (teaching, training, youth work, etc.) in both EU programme partners and Belarusian partners. Moreover, the exchange of experiences and good practices among partners will increase capacities of all participating institutions and other stakeholders interested in the modernisation and internationalisation of HE (e.g. other Belarusian and EU nonparticipating HEIs). The project aims, through the planned collecting and data analysis from surveys, the planned study visits, the information collected about good practices cases in defining learning outcomes, the way of the transition shift from using traditional input/content to output/outcomes approaches in Belarusian participating universities, the implementation of student-centred teaching models, the updating and integrating teaching methodologies and changes in HEIs management, and the recognition and validation of competencies, will be not only used in Belarus participating universities, but will also be further disseminated at public events (dissemination events of the project).

Different strategies and dissemination actions of FOSTERC project results are used to achieve the expected impact and giving continuity to the project towards other HEIs, both in EU programmes countries and regions where there is a potential that such know-how could be used, especially towards 'Convergence' regions in Central Europe which aspire to promote the modernisation and internalization of their HE improving quality of their institutions. The project focuses on key target groups at different levels:

INTERNAL LEVEL, where is envisaged that the partners disseminate findings to other colleagues, their regional and national contacts (e.g. public authorities, project partners, etc.) (also in national languages), highlighting the contribution of the FOSTERC project to society.

REGIONAL and LOCAL LEVEL, where partners of the FOSTERC project are committed to reach efficiency and effectiveness in communication, by using all the opportunities linked to the project's field of action and to communicate the project results in their own language (e.g. regional and local institutions, media, associations, civil society, private sector actors), raising awareness of the topic, informing a wide public, engaging different stakeholders and promoting the project as a good practice.

EUROPEAN and INTERNATIONAL LEVEL, where the impact of the project takes the form of networking with other relevant countries. FOSTERC partners are actively involved in international networks that will enable efficient and thorough exchange of FOSTERC findings with the results of other relevant projects.

Each of these groups (e.g. European Institutions, Organizations, Associations, Foundations, etc.) contribute to a social learning process in which the sharing of experiences and knowledge jointly contributes to a shared understanding.

V. CONCLUSION

FOSTERC project is intrinsically sustainable because it is appropriate within its broad overall objective, addressing national priorities and the evolving needs within the participating stakeholders, providing solutions that will be replicable across other Belarusian non-participating HEIs and other non-partner countries. The consortium covers HEIs from 4 EU countries (Spain, Lithuania, Poland and Finland) and 8 Belarusian HEIs from different geographical areas representing well the Belarusian context. Structural barriers (political, economic, social, cultural) could restrict the occurrence and impacts of the FOSTERC project. However, the wide coverage of the consortium will mitigate these barriers. Moreover, the project has as a full partner the Ministry of Education and the National Institute for HE, both will encourage more for a discussion of a broad range of policy makers and other stakeholders in the policy making process integrating FOSTERC's evidence into policy and reducing possible barriers.

Staff of Belarusian partners will attend training activities at home and EU visits. These capacity building processes will have a long-term impact on the skills of academic and administrative staff. Moreover, the support and follow-up of the EU partners to Belarusian participating universities in the way of description of learning outcomes from the beginning will guarantee that at the end of the

project, Belarusian participating HEIs will be able to develop full pedagogical models based on competencies. These approaches could be replicated in any other non-partner HEI and it is quite likely that the culture of learning outcomes based on competencies development expands for the whole Belarusian HE system.

Finally, the FOSTERC consortium will interact with worldwide research groups and forums to set up cross-fertilization of ideas and concepts and together give continuity and extent the scope of the project and to establish a permanent network of FOSTERC.

REFERENCES

- [1] V. I. Baydenko, Competences: to development of competence-based approach, Works of a methodological seminar "Russia in Bologna Process: problems, tasks, prospects". Moscow, Research center of problems of quality of training of specialists, 2004, pp. 25-30.
- [2] V. I. Baydenko, Competences of professional education (to development of competence-based approach), Higher education in Russia, (11), 2004, pp. 17-22.
- [3] V. A. Bolotov, V. V. Serikov, Competence-based model: from the idea to the educational program, Pedagogics, (10), 2003, pp. 26.
- [4] Competence-based approach in pedagogical education, SPb, RSPU publishing house, 2004.
- [5] N. F. Efremova, Forming and estimation of competences of education, Rostov-on-Don, Arkol, 2010.
- [6] J. Hartig, E. Klieme, D. Leutner, Assessment of competencies in educational contexts: State of the art and future prospects. Göttingen: Hogrefe & Huber, 2008.
- [7] O. E. Lebedev, Competence-based approach in education. School technologies, (5), 2004, pp. 3-12.
- [8] D. C. McClelland, Testing for competence rather than testing for "intelligence". American Psychologist, (28), 1973, pp. 1-14.
- [9] R. J. Shavelson, On the measurement of competency. Empirical Research in Vocational Education and Training, 2 (1), 2010, pp. 43-65.
- [10] I. L. Shishov, Competence-based approach to education as need. World educations — education in the world, (4) 2005, pp. 41-43.
- [11] A. K. Wigdor, Performance assessment for the workplace (Vol. I). Washington, DC: National Academy Press, 1991.
- [12] I.A. Winter, Core competencies – a new paradigm of result of education. Higher education today, (11), 2005.

Information technology for video broadcasting

I. Olenchenko, A. Oliinyk, S. Subbotin

Abstract—This paper presents description of real-time video broadcasting features, created business process diagrams that shows main functions of developed software, analysis of the client-server architecture, review of protocol RTMP and exiting multimedia server software such as Wowza Streaming Engine. Basing on the results of this analysis software tools for system development were selected.

Keywords— Android, client-server, geolocation, real-time video broadcasting, RTMP, Wowza Streaming Engine.

I. INTRODUCTION

Real-time video broadcasts are rapidly gaining popularity because there are a lot of different events and incidents in the world and around us.

Online video stream is a "compressed" video and audio information transmitted over the Internet in real-time. Information technology advancement make possible to broadcast not only from special equipment and personal computers but also through smartphones. Also online broadcasts are being implemented in social networks and video platforms.

The area of video broadcasting covers many spheres of human life and work and nowadays it is very important to organize the broadcasting process from places of accident. This allows you to observe extraordinary events from the first person, respond quickly to these events and analyze what's happening. All well-known social networks and video platforms are also trying to implement live video broadcasting.

II. STATEMENT OF THE PROBLEM

The relevance of the project lies in the fact that despite the large number of existing mobile applications, not all have a well-implemented main functionality for video broadcasts and user-friendly and understandable user interface. These facts leads the development of software that will combine a user-friendly interface and a well-implemented main functionality.

The main goal of the project is development of software that will allow creating video broadcasts, viewing already created ones, organizing work with geolocation and a map. In addition, the software should support the registration and authorization of users.

In the context of the tasks, it became necessary to choose the most appropriate application architecture and study the existing methods of streaming video and audio signals.

III. ANALYSIS OF APPLICATION ARCHITECTURE

Creating a client-server architecture marked a new stage in development of network information technology. It has been made possible by increased internal and external memory, computer speed, data rates and the advent of powerful mobile devices - smartphones and tablets.

The main idea of the client-server concept is to deploy servers on powerful computers and client applications - on less powerful computers or on mobile devices such as smartphones and tablets. Due to this, resources of a more powerful server and less powerful equipment of clients will be involved.

I. Olenchenko Zaporizhzhya National Technical University, Zaporizhzhya, Ukraine (e-mail: igor olenchenko.232@gmail.com).
A. Oliinyk Zaporizhzhya National Technical University, Zaporizhzhya, Ukraine (e-mail: olejnikaa@gmail.com).
S. Subbotin Zaporizhzhya National Technical University, Zaporizhzhya, Ukraine (e-mail: subbotin.csit@gmail.com).

Therefore, "client-server" architecture was chosen as the basis for the implementation of future software.

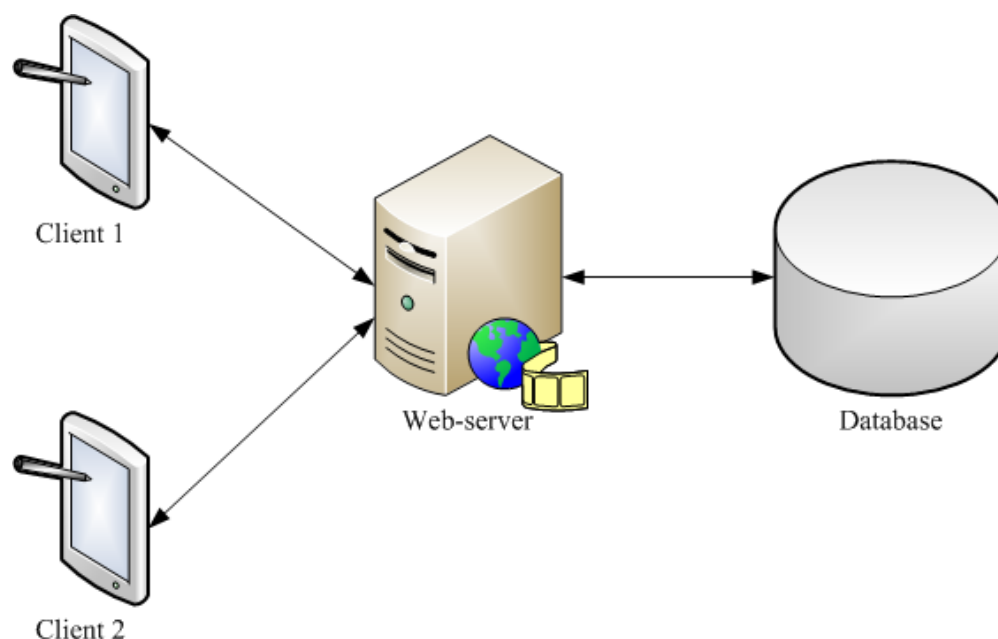


Fig. 1 – Overall structure of developed software using client-server architecture

The concept of "client-server" made possible to combine the positive qualities of systems with one user (high level of dialogue support, user-friendly interface) with the advantages of large computer systems (integrity support, multitasking) [1].

Basic functional was analyzed and allocated using concepts of client-server architecture that shown in fig.2. For the client side main task is implementation of user interface and work with camera of device. For the server side main task is processing data that comes from the client and saving it to a server.

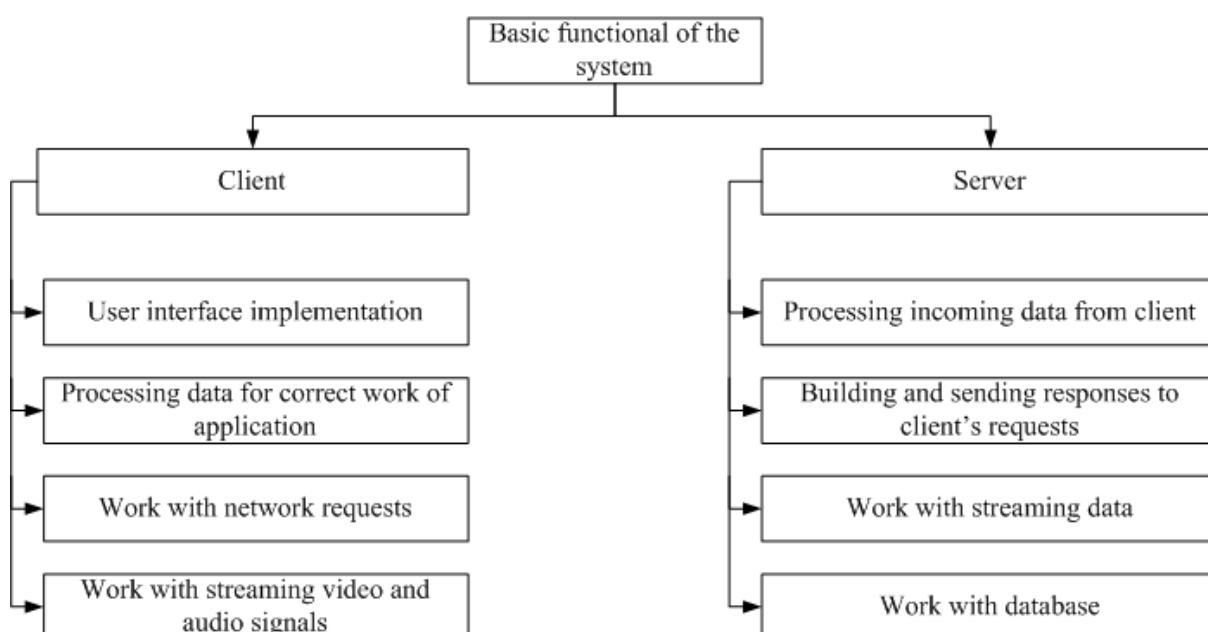


Fig. 2 – Overall structure of basic functional of the system

IV. REALIZATION OF INFORMATION TECHNOLOGY FOR VIDEO BROADCASTING

During the design phase of the system, information models were created in the form of business process diagrams. Their features consists in fact that with help of a certain set of symbols you can present any process in the form of a block diagram [2]. In our case created diagrams describe the main software functions such as user registration, creation and viewing streams. The diagrams are shown in Figures 3-5

Figure 3 shows a diagram, which describe process of user registration. It shows possible user actions during registration and reaction of the system to incorrect user actions - input of incorrect data and shows interaction of the client and server parts of the software.

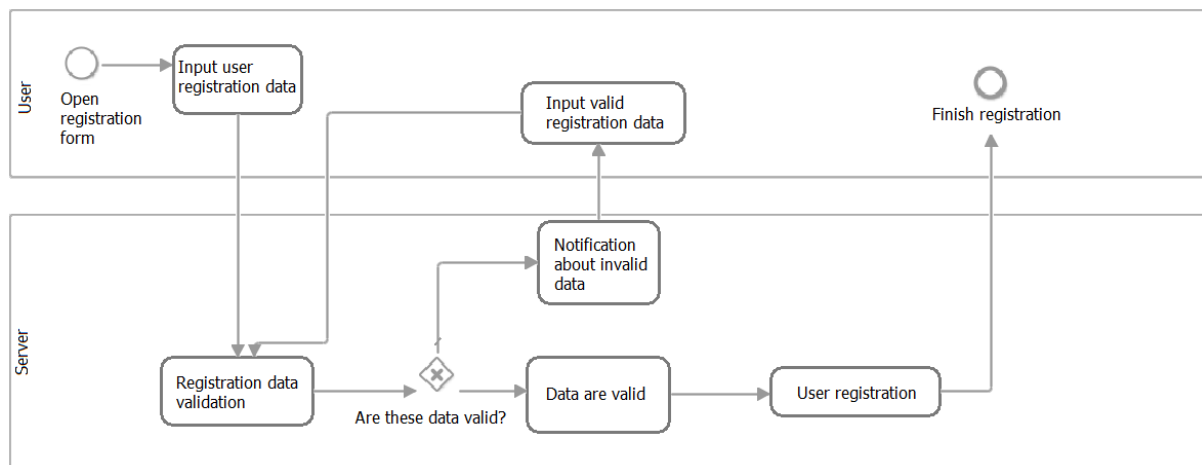


Fig. 3 Diagram “User registration process”

The figure 4 shows a diagram of the process of creating a broadcast, describing the basic actions needed to create the broadcast, the system's response to user actions and the interaction between client and server.

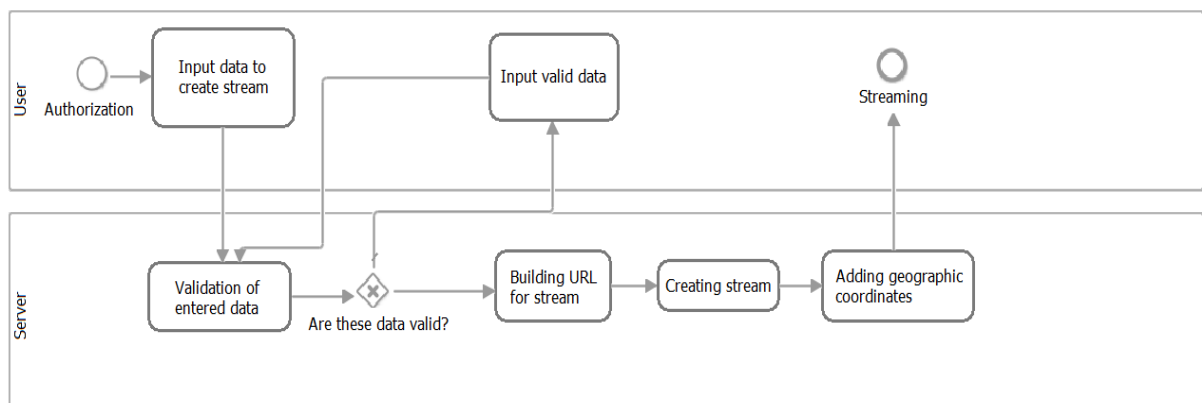


Fig. 4 Diagram “Creating stream process”

Figure 5 shows a diagram, which describe process of selecting and viewing already created broadcasts. The structural elements of this diagram are similar to those that used in the diagrams pictured in Figures 3 and 4.

The visualization of the interaction between the client and server parts of the software was an important component in design of the future system because it clearly demonstrated the advantage of client-server architecture over other types of architectures.

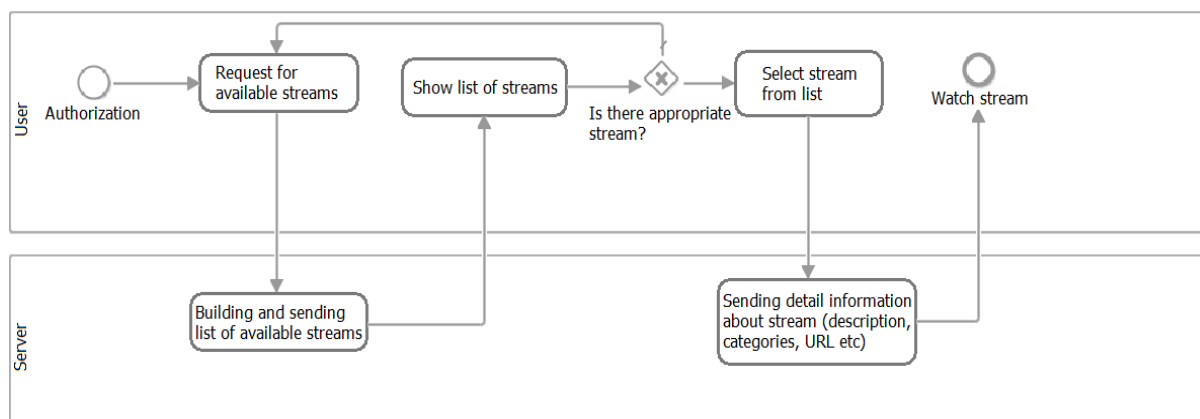


Fig. 5 Diagram “Watching stream process”

Based on the developed diagrams (Figures 3-5), main requirements for the software being developed were specified. In addition, created diagrams influenced the choice of system architecture.

It was necessary to divide the server part into a multimedia server and an HTTP server for accomplish this objective. The HTTP server was designed to receive and process data that comes from the client (information about user and streams) and organization of work with the database. HTTP server implemented via Node.js programming platform, JavaScript programming language [3-7], MongoDB has been chosen as database management system [8-10]. The interaction between the client and server parts was implemented using the developed API-interface [11-14].

Multimedia server was designed to work with streaming video and audio signals, for the organization of its work using the Wowza Streaming Engine software.

Wowza (full name - Wowza Streaming Engine) is a server software for streaming audio and video signals and video delivery on request. It has the ability to install on the following operating systems: Linux, Mac OS X, Solaris, Unix, and Windows [15]. RTMP is using for streaming video and audio.

RTMP (Real Time Messaging Protocol) is a streaming data protocol optimized for networks with weak bandwidth, mainly used to transfer streaming video and audio signals from cameras over the Internet [16].

The client part has been developed for Android operating system using Java programming language [17-26].

The main task that had to be solved in the process of creating a mobile application is the organization of video broadcasts using the built-in camera. To do this, it was necessary to avoid the use of the standard application "Camera", which is available for use in the development of applications and work directly with the camera of the smartphone. This allowed you to adjust the video quality by changing the resolution, working with the focus. After the connection to the server is established, the transfer of video from the camera to the multimedia server starts in real-time. Figure 6 shows the process of broadcasting video to multimedia server.

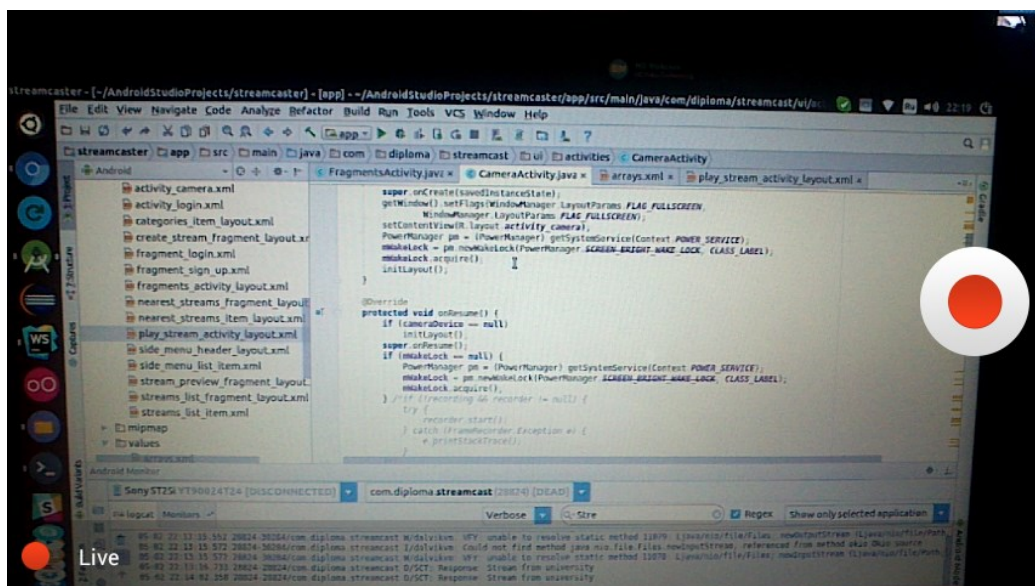


Fig. 6 Screen “Camera”

In addition, one of the features of the developed mobile application is the work with geolocation and map. While each stream is being created, geographic coordinates of the user are sent to the server, which created this broadcast. This allows you to display location of this stream on map with marker and serves to locate address where the broadcast will take place. (Fig.7).

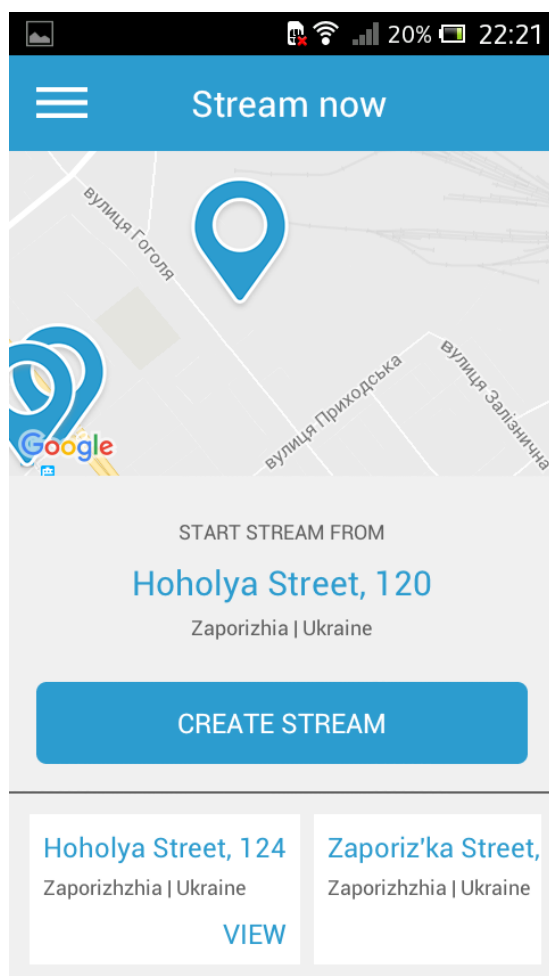


Fig. 7 Main screen of application

Geographic coordinates are used to sort the list of all streams. All broadcasts are sorted in descending order from the location from the user (Fig. 5). Calculation of the distance between position of user and location of the video is performed on the server.

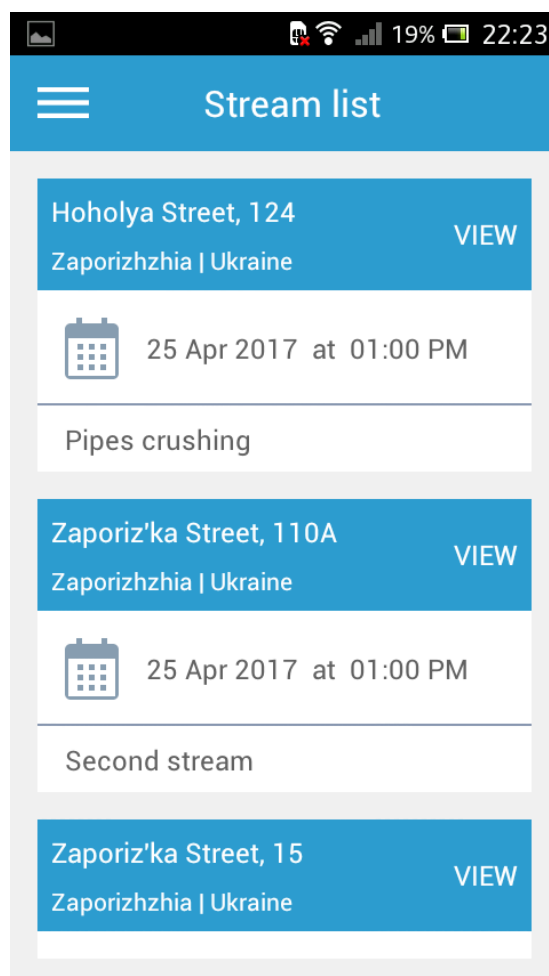


Fig. 8 Screen “Stream list”

These features of the user interface simplify the user's work with the application [27]. The usability of the user interface and implemented functionality of the system make it possible to recommend the developed software for use in practice for work with real-time video broadcasts.

V. CONCLUSION

Thus, the following tasks have been solved:

1. The requirements for the system being developed are formulated.
2. Business process diagrams describing the main functions of the software and demonstrating examples of interactions between the client and server parts were developed.
3. The architecture of the application, which allows the most effective implementation of software for working with video broadcasts, was chosen.
4. The server part of the software, which allows processing and storing user data, information on video broadcasts and organizing the streaming of video and audio signals, is implemented.

5. The client part is implemented as a mobile application that runs under the operating system Android 4.0 and higher, has a user-friendly interface, maximally oriented to the majority of users.

The developed software is the first version of the project, in which the basic functions necessary for the normal functioning of the system are realized. In future, it is planned to expand the basic functionality and work on user interface.

ACKNOWLEDGEMENTS

This publication is the result of the project implementation: TEMPUS CERES: Centers of Excellence for young REsearchers (Reg. no.544137-TEMPUS-1-2013-1-SK-TEMPUS-JPHES).

REFERENCES

- [1] Client-server architecture. URL: <http://buklib.net/books/23148/>.
- [2] Business Process Model and Notation (BPMN) version 2.0 documentation. URL: <http://www.omg.org/spec/BPMN/2.0/>.
- [3] Flenagan D. JavaScript. Karmanijj spravocnik. Moskva: Izdatel'skij dom "Vil'jams", 2015. 320 p. (In Russian).
- [4] Rejsig D. JavaScript. Professional'nye prijomj programirovanija. SPb.: Piter, 2008. 76 p. (In Russian).
- [5] Herron D. Node.js Web Development, 3rd Edition, Packt Publishing, 2016. 376 p.
- [6] DelBono E. Node.js Succinctly. Syncfusion inc, 2016. 76 p.
- [7] Braun I. Veb-razrabotka s primeneniem Node i Express. Polnocennoe ispol'zovanie steka Javascript. SPb.: Piter, 2017. 336 p. (In Russian).
- [8] Fowler M., Sadalage P.J. NoSQL Distilled. A Brief Guide to the Emerging World of Polyglot Persistence. Addison-Wesley, 2013. 171 p.
- [9] Ward H., Latino J. Pentaho Analytics for MongoDB Cookbook. Packt Publishing, 2016. 250 p.
- [10] MongoDB documentation. URL: <https://docs.mongodb.com/manual/>.
- [11] Pereira C. R. Building APIs with Node.js. Apress, 2017. 142 p.
- [12] Wandschneider M. Learning Node.js: A Hands-On Guide to Building Web Applications in javascript, 2nd Edition. Addison-Wesley Professional, 2016. 320 p.
- [13] Patni S. Pro RESTful APIs: Design, Build and Integrate with REST, JSON, XML and JAX-RS. Apress, 2017. 126 p.
- [14] Bojinov V. RESTful Web API Design with Node.js. Packt Publishing, 2015. 138 p.
- [15] Wowza Streaming Engine documentation. URL: <https://www.wowza.com/docs/>.
- [16] Real Time Messaging Protocol (RTMP) specification. URL: <http://www.adobe.com/devnet/rtmp.html>.
- [17] Ye. R. Android System Programming. Packt Publishing, 2017. 470 p.
- [18] Muhammad A. OpenCV Android Programming by Example. Packt Publishing, 2016. 171 p.
- [19] Kousen K. Gradle Recipes for Android: Master the New Build System for Android. O'Reilly, 2016. 168 p.
- [20] Darwin I. F. Android Cookbook: Problems and Solutions for Android Developers, 2nd Edition. O'Reilly, 2017. 838 p.
- [21] Horton J., Vasconcelos H., Portales R. Android: Programming for Developers. Packt Publishing, 2016. 1136 p.
- [22] G. O. Young, "Synthetic structure of industrial plastics (Book style with paper title and editor)," in *Plastics*, 2nd ed. vol. 3, J. Peters, Ed. New York: McGraw-Hill, 1964, pp. 15–64.
- [23] Shildt G. Nouton P. Java 2. Naibolee polnoe rukovodstvo, 3-e izd., pererab. i dop. SPb.: BHV-Peterburg, 2007. 1034 p. (In Russian).
- [24] Friesen J. Java XML and JSON. Apress, 2016. 300 p.
- [25] Dejtjel P., Dejtjel H., Uold A. Android dlja razrabotchikov. 3-e izdanie. SPb.: Piter, 2016. 512 p. (In Russian).
- [26] Goloshchapov A. L. Google Android. Sistemnye komponenty i setevye kommunikacii. SPb.: BHV-Peterburg, 2012. 384 p. (In Russian).
- [27] Goloshchapov A. L. Google Android: programirovanie dlja mobil'nyh ustrojstv. 2-e izd., pererab. i dop. SPb.: BHV-Peterburg, 2012. 448 p. (In Russian).
- [28] Clifton I. G. Android User Interface Design: Implementing Material Design for Developers, 2nd Edition. Addison-Wesley Professional, 2015. 448 p.

Virtualization-based architecture for the operating systems history demonstration

Anatoliy V. Gusev, Vladimir Yu. Kovalenko, Dmitriy A. Kostiuk¹

Abstract—The universal virtualized infrastructure for embeddable demonstration of mobile and desktop operating systems is presented, with emphasis on live display of the graphical user interface history. Experience of using virtual machines instead of screenshots in a visual timeline of GUI is reviewed. Presented solution provides user control via the Internet or intranet web access, and QEMU-based nested virtualization for the unified interaction with heterogeneous simulated architecture of mobile devices. Availability of materials is considered, as far as approaches used to improve web-based user interaction with virtualized GUI. Internal architecture, deployment and scaling principles are discussed.

Keywords—virtualization, history, computer interfaces, live demonstration.

I. INTRODUCTION

Although technically working with the new information technologies doesn't demand knowing the history of their development, however specialist who formulates or applies modern theory without knowledge of its history, runs the risk of repeating the mistakes of predecessors personally one by one [1].

Being in the context of spoken above statement we have developed the osthline project [2], intended to provide the user (either via network or on a stand-alone workstation) with a set of chronologically-related HTML documents, each one with a description of the specific graphical operating system and its live illustration in the form of built-in frame with the screen of a running virtual machine (thanks to the performance of today's notebooks and desktop PCs, this task is easy enough), as shown on Fig. 1.

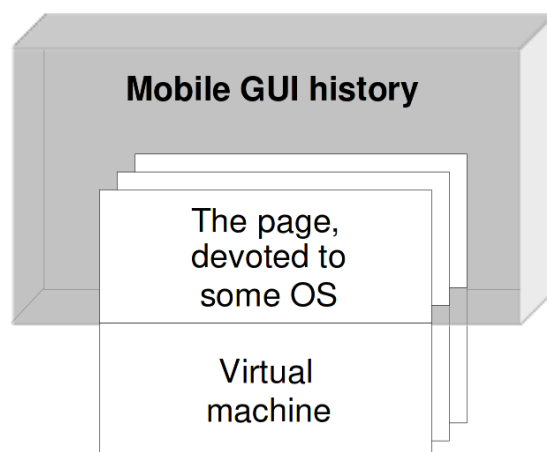


Fig. 1 Scheme of the virtualization-enhanced interactive document

Nowadays there is some amount of web based projects demonstrating historical operating systems with their 'screen' embedded into the Internet page. One of such projects is oldweb [3, 4] (dedicated to vintage web browsers), and there are several others (e.g. single demos of some old Microsoft or Apple OS, and Amiga). However we still didn't see web demonstrations dedicated to vintage mobile operating systems (perhaps due to some technical difficulties, discussed later). This fact increases the relevance of presented here project, based on the original university

¹ Anatoliy V. Gusev, Brest State Technical University, Brest, Belarus (e-mail: avg.tolik@gmail.com)
Vladimir Yu. Kovalenko, Brest State Technical University, Brest, Belarus (e-mail: vl.kovalenko1989@ya.ru)
Dmitriy A. Kostiuk, Brest State Technical University, Brest, Belarus (e-mail: dmitriykostiuk@gmail.com)

Authors are grateful to CERES project for the financial support

course of the graphical user interface history, which itself includes 30 mobile operating systems and 40 desktop ones, user interfaces of which turned out to be a historical milestone.

II. TECHNICAL INFRASTRUCTURE OF THE HTML-BASED VIRTUAL MACHINES DEMONSTRATION

Client-server applications with web interface have gained today the widest usage, and are already replacing classic desktop applications in some areas. List of reasons backing such move includes the presence of a web browser on all platforms and architectures, as well as sufficient performance of a JavaScript client-side web applications are built on (at least in modern browsers). Universal accessibility (including platforms with a touchscreen interface and other so-called 'thin clients', i. e. on mobile and portable devices, targeted mainly at the use of cloud services) makes browser the most convenient entry point for the end user application / service. This is especially significant when the main computational load is placed on the server and / or other nodes in the network, and the client is used only for management and access to resources (the reverse case, unfortunately, at this moment is not effective as far as JavaScript implementation involves larger overhead costs compared to traditional languages on server platforms).

One of the tasks that can get significant benefit in case of the client-server implementation based on a private cloud, is virtualized applications exhibition, either in demonstration purposes or for some usability comparison & research, etc.

Problem of demonstrating software in case of strong fragmentation in target hardware platforms of correspondent devices is rather obvious. Therefore developers of such exhibitions have to use set of emulators with disk images containing a variety of mobile operating systems to run. In this case, even if the regular tools of desktop virtualization would provide a convenient interface to simultaneously run multiple instances of different emulators, low productivity of the emulation associated with different processor architectures would cause negative impact on the productivity and efficiency (especially noticeable during boot process of the emulated hardware). Also it turns out sometimes that local emulation do not fit the workflow at all: sometimes a shared access to a copy of the test application is required, as well as a demo access [5].

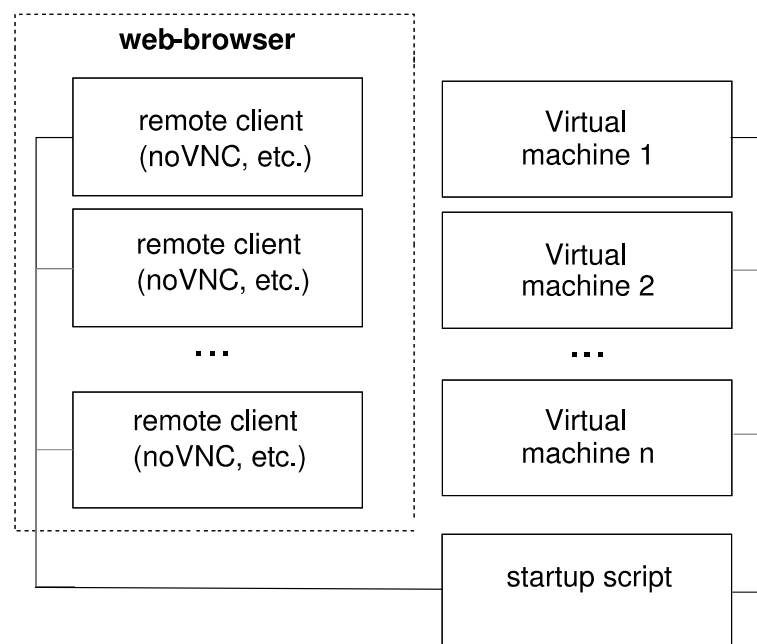


Fig. 2 General scheme of the virtual machines interaction

Technical infrastructure we have proposed to implement such demo materials, is briefly reviewed in [2] and includes the following components:

- QEMU virtual machine,
- some remote access client written in JavaScript and HTML5 (e.g. noVNC),
- JavaScript framework to display information in one interactive timeline, or sequence of slides, etc.

Running the scheme shown in Fig. 2 is done by a startup script. This script scans subdirectories in search for the information elements: pages with the content, virtual machine images and scripts to run them. Regardless of used remote access protocol (VNC or some other), script passes port numbers to discovered virtual machines and reconstructs HTML document to include pages with information materials into the demo timeline. This component-based approach allows to divide the information, making some parts of the exposition publicly available on the network, and taking away from public access those virtualized systems which can not be redistributed due to the terms of commercial licenses.

III. EMULATORS OF THE MOBILE AND DESKTOP HARDWARE

Virtual machine is used as a fully isolated container storing a snapshot of the already running OS. Choice of QEMU is caused by the extremely simple transfer of virtual machine images between computers, and also by its ability to emulate not only x86-compatible platforms, but also ones based on ARM, MIPS, and Motorola 68k processors, which are mandatory for some vintage operating systems.

However, currently multiplatform capabilities of QEMU are poorly used: support of motherboards and peripheral devices on alternative platforms at this emulator is not always sufficient, especially in case of really old operating systems.

At the same time, there are emulators available to support virtually all vintage Intel-incompatible systems – as community-developed emulators, so ones created as the part of some commercial SDK. The first variant is more common in desktop operating systems segment (they are used to emulate Xerox Alto, Apple Lisa and Mac computers, Amiga, etc.). For now oimeline uses only one community-developed mobile OS emulator – Open Einstein project, which allows to run NewtonOS.

As opposed to them, emulators from proprietary mobile devices SDK are not typical for desktop segment but are dominating in mobile one, such as Psion EPOC16 and EPOC32, PalmOS, Magic Cap, Windows CE operating systems, and pre-release versions of Android SDK.

Unfortunately, all these specialized emulators don't support snapshots and remote access. Therefore a large part of demos is build along with the the nested virtualization scheme (see Fig. 3), where QEMU plays the role of an external container.

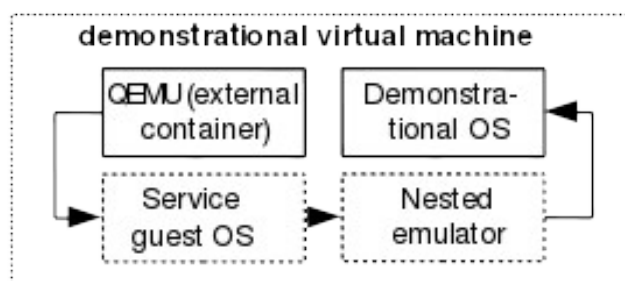


Fig. 3. Nested virtualization of the mobile OS

Of course substantial part of operating systems do not require nested virtualization and so internal emulator is not used for them. That's the case of different versions of Windows, OS/2, GNU/Linux Maemo, Android, WebOS (not least due to the fact that QEMU is often included in mobile SDK) and some others.

There is one more important component on Fig. 3, a service guest OS, which is used to start the nested emulator. Its choice is determined by the following three requirements: minimal memory consumption, usage of idle CPU cycles, and, preferable, support for USB bus emulation, which allows to emulate pointing devices in absolute coordinates (their usage will be discussed later). We have used FreeDOS and ReactOS as a service guest OS in addition to multiple versions of GNU/Linux. It should be further noted that ReactOS (a small Windows-compatible OS with an open source license) perfectly meets all three requirements, and thus in our own experience it is the first case of its successful practical usage.

IV. AVAILABILITY AND THE LEGAL ASPECTS OF RUNNING HISTORICAL OS

Noticeable part of historically-significant operating systems are now a free/libre software (ones having such licenses from the very beginning, ‘opensource’ because of the extreme aging or being a free clone of the abandoned commercial system). That’s about different DOS- and Linux-based environments for desktop and mobile computers (such as GEM, CDE, Qtopia, Gnome Palmtop Environment, OpenMoko, Maemo, Android, etc.). So this part of the timeline is freely distributable (e.g. for educational and/or research purposes).

The timeline material is still at the stage of filling, and at this point the review has some missing objects, which have played an important role in the history of interfaces. But the number of historically important GUI shells without runnable versions appears to be surprisingly small: e.g., this row currently includes only two mobile systems: PenPoint OS and IBM Simon.

V. MULTI-USER INFRASTRUCTURE OF THE VIRTUAL MACHINES MANAGEMENT

Fig. 2 shows simplified architecture capable to provide user with the demonstration containing number of liveGUI demos. Such approach can be sufficient in case of relatively small amount of virtualized systems and local single-user access. But more complex solution is unavoidable in case of Internet or intranet resource simultaneously accessed with a number of users. General structure of such server is shown on Fig. 4.

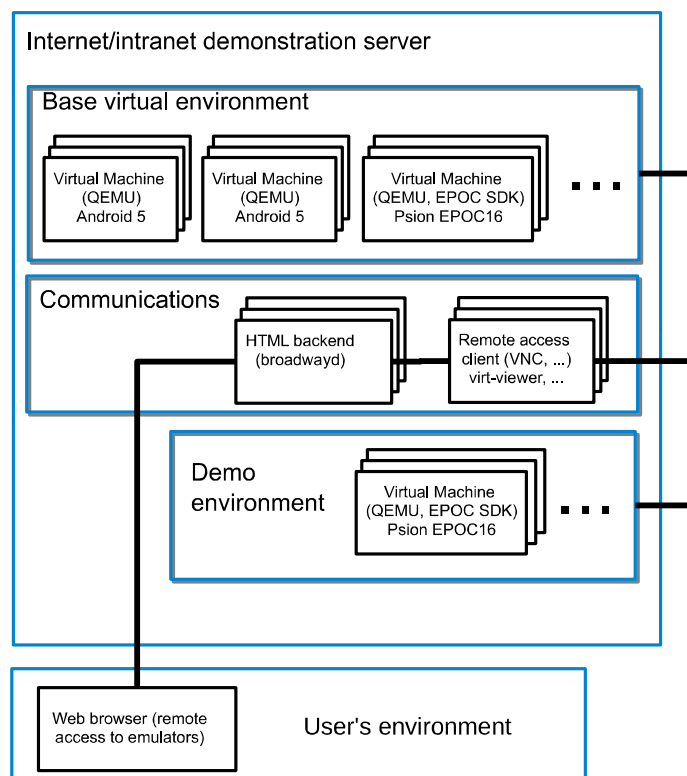


Fig. 4. The structure of the Demo server

As in case of the traditional embedded development tools stack, the demonstrated OS is on various types of emulated devices, which can be nested inside of an outer virtual machine for the convenience of the deployment and usage.

This approach significantly simplifies management of the heterogeneous set of the emulated systems, and retains all the benefits and convenience provided by server virtualization of desktop systems as far as possibilities of the automation, cloning and integration into the enterprise infrastructure, if needed.

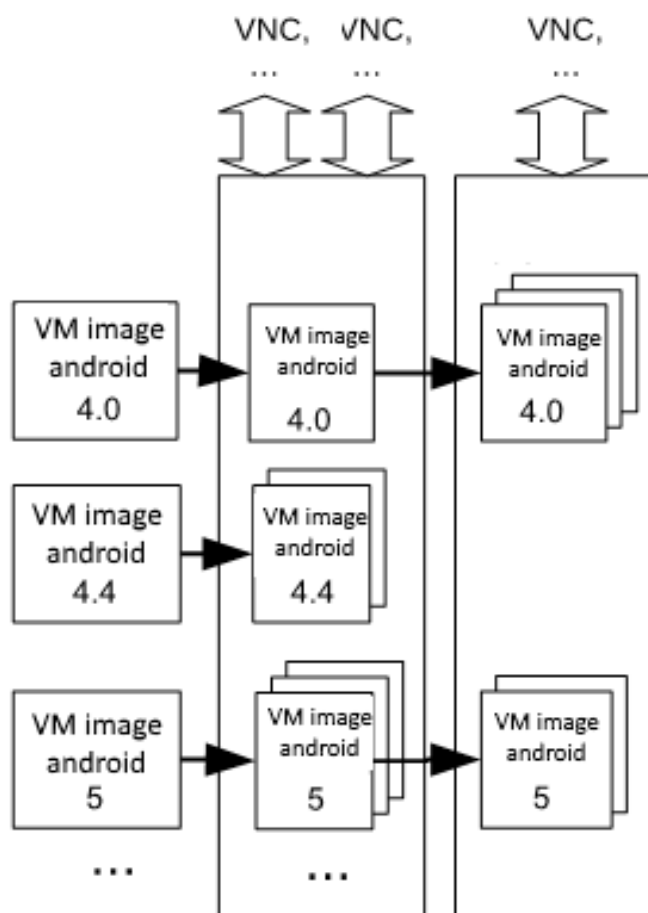


Fig. 5. Virtual machines images cloning example

Server contains reference images of a nested VMs with different OS (Android of various versions on Fig. 5). These images are configured and tuned for the fastest possible launch of a new instance. When user needs to interact with a personal exemplar of the virtualized environment, one of the reference images cloned to start the demo.

Technically image management is provided by a set of rather simple scripts that perform on-demand cloning of master images, as well as start and stop of the VM execution.

VI. MORE ON WEB ACCESS

Users are accessing the desired VM through the web page. Initially we have used noVNC remote client based on JavaScript to access an emulator via the web environment, as described in [2]. However, in the process of testing some advantages were discovered for the access via the broadway GTK backend – a rendering subsystem in HTML 5 embedded in the latest version of the GTK library. The structure of this web-based access solution built into the server is shown on Fig. 6.

On the client-side access to the emulator is done via the web browser, which communicates with the broadway GTK backend, and the broadway draws the window of the chosen GTK3

applications on a HTML5 canvas object. The graphics program that uses GTK library runs in headless mode, i.e. only for network access. System daemons `broadwayd` (each on its own single network port) should be run for this, and the graphics program should be supplied with an additional set of variables at startup, which are read by the GTK library and thus determine the rendering mode. As a result, the startup code of the `broadway` system daemon inside of the starting script looks as follows:

```
broadwayd :1 &
```

```
GDK_BACKEND=broadway UBUNTU_MENUPROXY= LIBOVERLAY_SCROLLBAR=0  
BROADWAY_DISPLAY=:1 VNCViewer
```

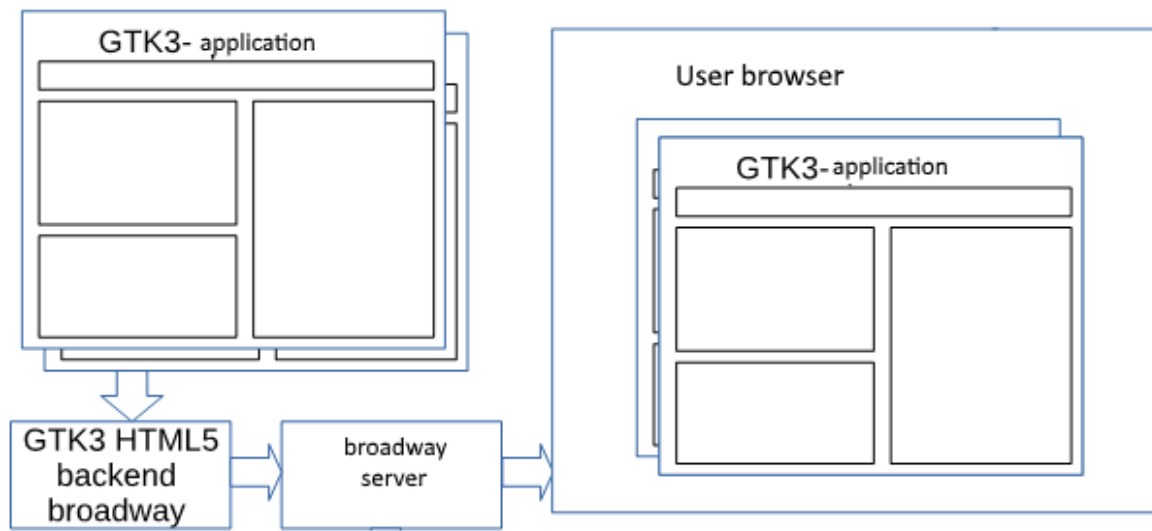


Fig. 6. The Broadway operation principle

One of VNC or SPICE clients written using GTK3 plays the role of the graphics program running inside of the external VM and connected to the embedded VM (e.g. VNCViewer client can be used as far as QEMU graphical window); Broadway technology, in its turn, translates the user session access to the browser outside of the server.

The choice of this web access architecture is influenced by two considerations: the best performance of the code (which will be discussed below), and no need for additional components of the JavaScript based VNC (broadcasting a TCP connection to the web sockets) [5]. As for additional components on the client side, which would be necessary when using the VNC-client in JavaScript, embedded in HTML page, their necessity is dictated by the fact that the code in JavaScript (executed by the web browser) does not have access to the TCP protocol used for remote access, and therefore requires the application proxy running on the client side and transmitting TCP traffic to a web sockets.

VII. NOVNC AND BROADWAYD PERFORMANCE TESTING

Obviously, the reason for better performance of the selected web access solution in comparison with JavaScript VNC client is caused by a highly specialized web server that is a part of Broadway, and the VNC-client running inside the external VM (both components are written in C). However, this code is executed on the server, while the use of VNC-client in JavaScript increases the processing load on the client machine. In turn, Broadway also connects to the client code in JavaScript, executed by a Web browser for the reception, transmission, and rendering.

To assess the real broadband impact on the performance of the developed solution, performance testing was separately carried out for the web access (configured on the separate system with Ubuntu 15.04 GNU/Linux OS, the AMD FX-8320 processor and the RAM capacity of 16 GB). The OS choice is subject to the presence of the components required for broadband operation (GTK library version 3.8 and above, as well as some software using it). The client and server parts of the system were running simultaneously on the server, providing equal impact of the relatively high resource consumption of the selected OS GUI for both sides, so it did not affect the comparison.

The noVNC JavaScript client was used for comparison. Load was assessed by the «load average» parameter, produced with the help of htop tool, which is the most common and generally accepted indicator to assess the server workload.

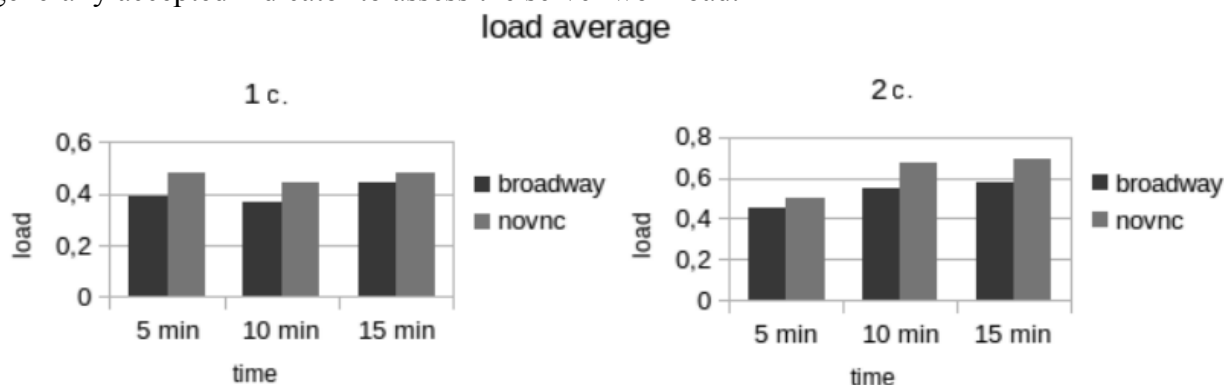


Fig. 7 noVNC and broadband comparison results

The comparison results are shown in Fig. 7. The analysis shows broadband advantage in terms of load, and this difference increases at growing number of connected computers (user's workplaces) to essential numbers when the number of virtual desktops reaches tens or hundreds.

Lesser server load of broadband, compared to VNC protocol, can be attributed to several factors: the lack of need for a separate VNC server, absence of operations associated with compression and stream encoding for VNC, more optimal and specialized server code. RAM consumption in the case of broadband allows to save 30-40 MB per virtual workplace.

As for the load on the client machine and the amount of generated traffic, both approaches are demonstrating parity (and both quantities are negligible). However, there broadband advantage is a complete lack of image artifacts that can occur when using VNC.

VIII. SOLVING THE PROBLEM OF NON-COINCIDING CURSORS

Running vintage OS in emulator often comes with a problem of non-coinciding cursors. This problem arises from the availability of two different types of cursor positioning devices: ones with relative coordinates (computer mice) and ones with absolute coordinates (tablets and touchscreens). Relative positioning devices provide OS with a vector of the cursor movement instead of it's new coordinates, and as a result, different cursor acceleration formulas are breaking coincidence of host and guest systems cursors [6].

Mainstream virtualization systems (including QEMU) can emulate as relative pointing devices (PS/2 and serial mice), so absolute pointing ones (USB Wacom tablet in case of QEMU). Mode with absolute coordinates allows host cursor to control guest OS and is called «mouse integration mode» in desktop virtualization systems. Mouse integration is available only if guest OS has special driver from the virtual machine vendor, or supports USB tablet. In all other situations desktop virtual machines use mouse lock mode, when host cursor is hidden until some dedicated hotkey is pressed, and user can interact with the guest system only. This lock mode is not only less convenient for operation, but it is unusable for the web access to a virtualized system. So

projects providing web demonstration of old operating systems have to show two cursors moving with different speed (Fig. 8).

In our case guest systems with nested virtualization avoid this problem if ReactOS or GNU/Linux service OS works inside of QEMU (as far as these systems have driver for the USB tablet). Unfortunately, there are a lot of systems which need less convenient service OS (i.e. FreeDOS for the EPOC 16 emulator or Windows 9.x/MacOS 7.x of the emulator of Magic CAP). Also there are old mobile systems directly emulated by QEMU, such as Pen-driven version of Windows 3.x.

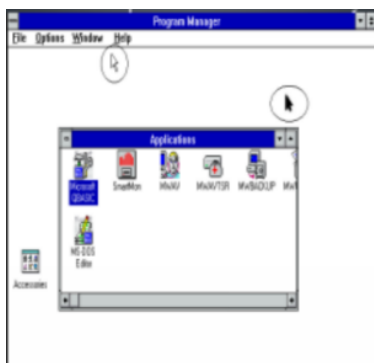


Fig. 8. Cursors non-coincidence caused by relative coordinates

Currently we have solved non-coinciding cursors problem for pre-USB service operating systems by creating a patch to QEMU, which implements emulation of the RS-232 Wacom tablet as an additional serial tablet backend. This sub-project was accepted for the Google Summer of Code program and have already landed in the mainstream QEMU version.

REFERENCES

- [1] K. Thibodeau, "Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years" in The State of Digital Preservation: An International Perspective, Conference proceedings, Washington D.C., April 24-25, 2002. Washington, DC, 2002, <https://www.clir.org/pubs/reports/pub107/thibodeau.html>
- [2] D. Kostiuk, P. Lutsiuk, S. Vlasenko, V. Zheludok, "Virtualization-based illustrated reviews of the software history." in Proc. of LVEE 2014 conference, Brest, "Alternativa publ.", 2014, P. 98-101.
- [3] A.J. Dellinger, "Oldweb.today lets you browse the Internet like it's 1999", The Daily Dot, Feb. 25, 2015. <https://www.dailydot.com/debug/oldweb-today-legacy-browser-simulator>
- [4] I. Kreymer, D. Espenschied, "oldweb.today: Web Archiving + Emulation", <http://labs.rhizome.org/iipc2016/owt.html>
- [5] V.Yu. Kovalenko, D.A. Kostiuk, "Virtualized web based farm for testing and demonstration of android applications with advanced security and portability", in Proc. of TIM-2016 conference, Grodno, 2016, <http://goo.gl/4VKups>
- [6] Д. А. Костюк, "Особенности использования виртуализованных окружений, внедренных в презентационные материалы" in proc. of OSSEDCONF-2012 conference, Alt Linux, Moscow, 2012, P. 83-86.
- [7] А.В. Гусев, Д.А. Костюк, "Решение проблемы интеграции курсора при запуске старинных операционных систем в QEMU" in proc. of OSSDEVCONF-2016 conference, Alt Linux, Moscow, 2016, P. 28-33.

Adaptive Algorithm of Intersection Management Based on Road Marking Adjustment Optimization

Aliaksandr A. Zavadsky, Vasiliy N. Shut

Abstract—The methods of transport flows managing has a number of shortcomings. An adaptive method based on the systematic configuration of the use of variable marking schemes of a road that leads to the junction, with the help of controlled signs is offered, which significantly increases its capacity and reduces the transport delay at the traffic lights. The paper proposes a formula presentation of markup and classifies the markings.

Keywords— adaptive control, road marking, formula presentation of markup, traffic light object, competing traffic flow.

I. INTRODUCTION

Modern road traffic is a complex dynamic system of people, vehicles and road infrastructure elements interference. To provide fast and safe traffic in such conditions requires the use of a complex of organizational and architectural planning activities for highways building, as well as the use of modern technical means of traffic organization based on information technology.

At the same time, the current experience in the use of technical means considers individual vehicles and does not take into consideration the interrelationships between the heterogeneous components of the road traffic [1, 2], which makes it urgent to develop new means of traffic automation (especially within the intersections) that allow its complex regulation based on constantly changing road conditions. The discrepancy between the existing street and road network of large cities in conditions of loaded and overloaded traffic flows makes the movement on regulated nodes more complicated. This situation requires reconsidering of the theory and practice of street traffic management and the use of modern tools and management methods.

Investigating traffic congestion as well as designing new transport management systems, special attention should be paid to such a road regulation element as the intersection. The intersection is the place of crossing of several roads. The most common crossing is X-shaped and it is the intersection of two roads. Here there is a decrease in the capacity of the way, since a part of time for the movement must be transferred to the intersecting (competing) stream. A traffic light at an intersection performs the function of a switching device that provides the alternate right for motion through a crossroads to motor vehicles (MV) from competing destinations.

Because of the stochastic change in MVs intensity approaching the intersection from different directions makes the task of satisfactory, adequate management of the crossroads difficult to achieve. All researches in the field of traffic management in the street network (SN) focuses on improving management at the SN nodes, through traffic control [3], i.e.: there is a resolution of the situation that has already taken place at the intersection. Frequently, the situation is very unfavorable with a large number of MVs stopped at the traffic lights. Control actions in this case are the changes in the traffic light cycle duration, time, the number of traffic lights and their order of succession [4, 5]. Often this set of changes in the parameters of a traffic light

A. A. Zavadski, Brest State Technical University, Brest, Belarus (e-mail: alexandr.zavadsky@gmail.com).
V. N. Shut, Brest State Technical University, Brest, Belarus (e-mail: lucking@mail.ru).

object (TLO) is not enough, and the "traffic jam" at the crossroads grows.

The failure of the functioning of the transport system occurs as a result of congestion on the transport network. The appearance of congestion, when there is a reserve of network capacity, is explained not only by the low optimality of traffic signaling, but also by its insufficiency. An additive change in road marking should be an addition to the traffic regulation at the crossroads by means of TLO.

The intensity of traffic in a city during the day, week and year varies significantly. Road accidents (traffic accidents), road closures, roadway repairs and other temporary factors also affect the change in the traffic intensity. Therefore, a significant change in the intensity of motion ratio along the directions is possible. In order to take into account the unevenness of the traffic, a flexible system is needed that allows changing the direction of motion along the lanes, for example, increasing the intensity of traffic to the left requires an increase in the number of lanes in a given direction. The system in this case changes the direction of motion along one of the bands with the help of controlled signs and the motion to the left switches to the mode of two-lane or one lane is used for motion right and left.

The ability to use variable schemas with managed signs and operational control increases the capacity of the SN significantly and reduces the transport delay, therefore requires extensive use in urban transport systems.

II. PROBLEM FORMULATION

The existing road marking has a number of shortcomings, which can be divided into two groups:

- Disadvantages for drivers.
- Disadvantages for the traffic police and municipal services.

One of the main problems faced by drivers in the process of driving on roads with standard marking is its permanence (immutability).

When driving along the stage to the next crossroads, the driver does not know the marking at this intersection until arriving. In this case, if the driver was not aware of prohibition to turn in a certain direction within a specific intersection, he may have to travel an additional distance to the next intersection, which involves additional time and material costs.

In the second group of deficiencies, one can mention an increased risk of road accidents. Thus, when moving along a stream (in a bundle) and approaching a crossroads in order to continue their movement at minimum costs and keep to the chosen route, drivers begin to rearrange themselves actively. Either if the subsequent marking is very different from the previous one or the car changes direction, or also in case of inattentive attitude of drivers to warning signs and sharp rearrangement, traffic accidents (road accidents) occur frequently.

In addition, the issues connected with the strength of standard markup in the process of exploitation appear. Thus, point 5.1.10 of General Specification 1231-2000 "Road markings: general characteristics" says that road markings made of thermoplastics, cold plastics or other similar materials should have functional strength of at least 1 year, and for paints and enamels the time-span should be not less than 6 months. [6].

However, a longer period is not established as well as the maximum lifetime, which causes the appearing of the places where marking is almost invisible or completely absent within the boundaries of cities. This lack, in addition, affects drivers who, without knowing anything, can violate the rules of the road.

Thus, the presence of a number of shortcomings that keeps from the effective implementation of traffic on urban roads requires the introduction of new means of regulation and control. It is proposed to install electronic scoreboards on the lanes between intersections that will choose the marking for vehicles moving in the stream and, if necessary, regulate the movement of cars

on the required tracks with minimal changes. At the same time, drivers are notified in advance about the form of marking at the upcoming crossroads, as the electronic scoreboard is installed at a distance of $\frac{3}{4}$ from the upcoming intersection, and the drivers have enough time for the necessary rearrangement.

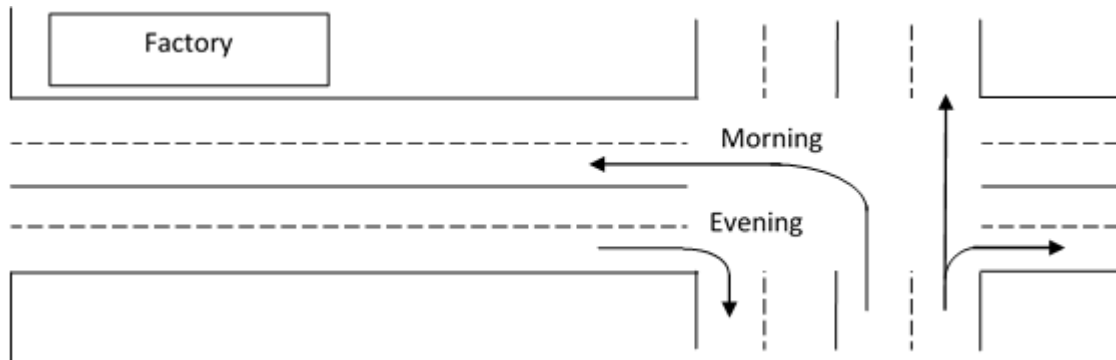


Fig.1 - SN fragment

Figures 1 and 2 show an example the road marking change during the day, regarding the structure of traffic flow in the morning and evening hours. At Figure 1 NS fragment with a clear center of gravity (factory) is shown. In the morning hours, MVs leave the main road at the crossroads to the left for the factory. In the morning, the markings shown are on Figure 2(a) or (b). In addition, marking 2(b) can pass more left cars. Although the marking 2(c) is typical for the evening. The indicator of marking change is the information from transport detectors located along each strip. For example, the appearance of a queue on the left lane before crossing, fixed by transport detectors, will be a signal for the control system to change the marking and increase the number of lanes for this category of transport.

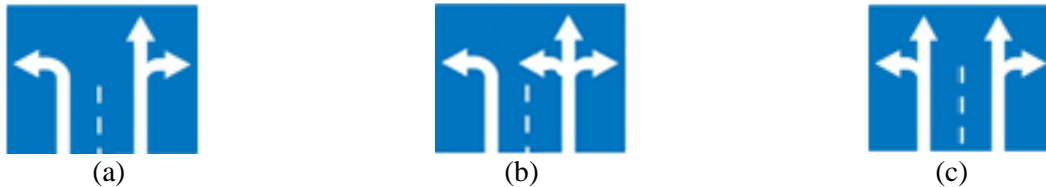


Fig.2 - Markup options for the represented NS fragment

III. FORMULA PRESENTATION OF MARKING

In order to manage the electronic scoreboard of the variable marking efficiently and hence the intersection, it is necessary to have a database of all possible markings. It is done for all the lanes of one direction of the road (the road is divided by a solid line into the forward and reverse direction). Each direction is divided by a dashed line into traffic lanes with the numbering $i = \overline{1, m}$ starting from the right edge of the curb of each direction of the road. The marking of the i -th lane is the vector \vec{P}_{ij} , pointing to one of the three directions $j = \overline{1, 3}$ of the traffic from the i -th intersection lane: to the right 1, straight 2 and to the left 3.

Therefore, the vector \vec{P}_{11} indicates the movement from the first lane to the right, and the vector \vec{P}_{12} for the movement along the first lane straight, the vector \vec{P}_{23} indicates the turn from the second lane to the left. Thus, the marking vector \vec{P}_{ij} reflects the possibility of moving in one direction or another with each of the lanes. Marking for one direction of the road is a collection of \vec{P} vectors \vec{P}_{ij} :

$$\vec{P} = \vec{P}_{11} + \dots + \vec{P}_{ij} + \dots + \vec{P}_{m3} \quad (1)$$

Where m is the number of lanes of one direction of the road, $= \overline{1, m}$, $j = \overline{1, 3}$. Here the sign "+" means not the arithmetic summation, but the union of the elements.

Figure 3(a) shows the intersection with the road markings for the forward direction of traffic. Figure 3(b) shows the road sign of this marking with its formula presentation. We arrange the brackets in the formula notation for this sign as follows: $\vec{P} = (\vec{P}_{11} + \vec{P}_{12}) + (\vec{P}_{22} + \vec{P}_{23})$. Then the set of vectors in each bracket refers to one lane of motion. This will be a formula presentation of a separate lane, that is, in each bracket, therefore, a set of vectors, or otherwise, directions of movement from a particular lane. Formulas of marking with banding along the lane are located in the second column of the Table I

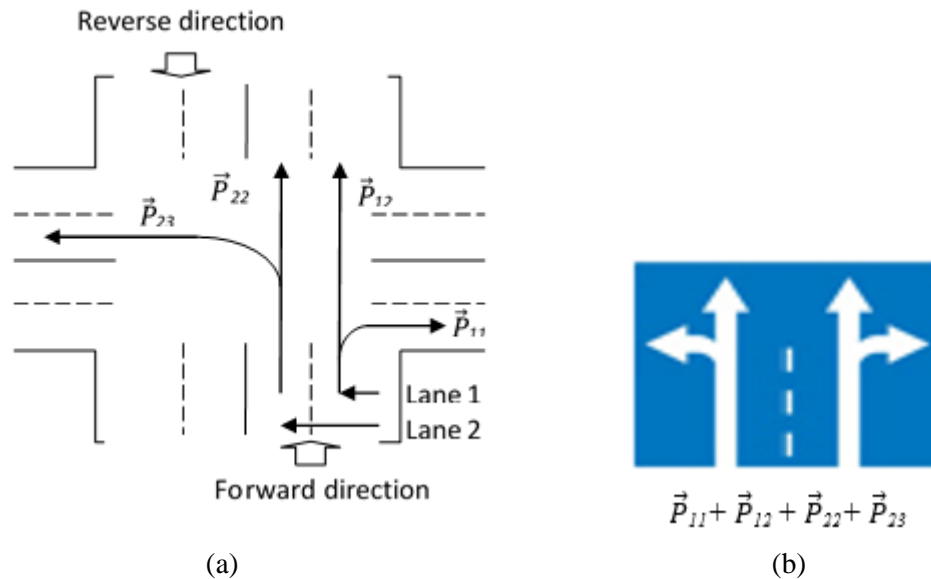


Fig.3 - Example of a two-way intersection and its corresponding sign with its formula presentation

It is possible to have a variant of grouping by the direction. If the first index (the number of the lanes) was decisive for grouping along the lane, then the second index of the vector (the direction of motion) is the determining factor for grouping along the direction. Arrange the brackets in the previous example as follows: $\vec{P} = \vec{P}_{11} + (\vec{P}_{12} + \vec{P}_{22}) + \vec{P}_{23}$. Record the markings when grouping in the direction we put in the third column of Table I.

Here the transformations in the form of summation of vectors are already possible. Vectors \vec{P}_{12} , \vec{P}_{22} have one direction 2 (forward). We replace the expression in parenthesis with the general direction vector forward \vec{P}_2 . In addition, we replace the vectors \vec{P}_{11} , \vec{P}_{23} by \vec{P}_1 and \vec{P}_3 . Then we obtain $\vec{P} = \vec{P}_1 + \vec{P}_2 + \vec{P}_3$. We denote by p the capacity of one lane (the number of cars passing through the cross section of the band per unit time, cars/min.). Then the modules of the vectors $|\vec{P}_1|$, $|\vec{P}_2|$ and $|\vec{P}_3|$ distributed as follows $0.5p$, p and $0.5p$ (see Figure 3). Coefficient 0.5 indicates that direction 1 (to the right), and direction 3 (to the left), in the given marking allocated 0.5(half) of lane. At the same time, a whole lane is allocated to direction 2 (forward).

We compose an ordered triple from the bandwidth of each direction ($|\vec{P}_1|$, $|\vec{P}_2|$, $|\vec{P}_3|$) or $(0.5, p, 0.5p)$. We normalize the triple by dividing each of its elements by $|\vec{P}| = 2p$ (the carrying capacity of the whole direction of motion along all its two lanes). Then we obtain $(|\vec{P}_1|/|\vec{P}|, |\vec{P}_2|/|\vec{P}|, |\vec{P}_3|/|\vec{P}|)$, or $(0.5p/2p, p/2p, 0.5p/2p)$, or $(0.25, 0.5, 0.25)$. The ordered triple $(0.25, 0.5, 0.25)$ is a numerical characteristic of a particular, given markup. Each coordinate of the triple (w_1, w_2, w_3) indicates the share w_j , $j = \overline{1, 3}$, which is the specific direction (right, forward, left) in the total capacity of the entire direction of travel. All the triplets are added to the fourth

column of Table I.

It should be noted that the recording in the direction is more informative, since it allows us to estimate the resource allocated by this marking for one or another direction of movement. The obtained numerical characteristics can be used for adaptive control by selecting the closest marking for an MV (bundled of MVs) approaching to the crossroads. The structure of the bundle (under the structure one understood the ratio of the number of cars traveling at the intersection to the right, forward and left) must coincide or be closest to the layout structure.

A pack \vec{X} of n vehicles $\vec{X} = \vec{X}_1 + \vec{X}_2 + \vec{X}_3$ is approaching to the intersection, where $|\vec{X}_1| = n_1$ is the number of cars that will go to the right, $|\vec{X}_2| = n_2$ is the number of cars going forward direction and $|\vec{X}_3| = n_3$ - to the left. Let us make the objective function of the situation from the marking parameter \vec{P} :

$$\min F(\vec{P}) = \|\vec{X} - \vec{P}\| = \|\vec{X}_1 - \vec{P}_1\| + \|\vec{X}_2 - \vec{P}_2\| + \|\vec{X}_3 - \vec{P}_3\| \quad (2)$$

Let us move from the parameter \vec{P} to the next objective function:

$$\min Z(\vec{P}) = |n_1/n - |\vec{P}_1|| + |n_2/n - |\vec{P}_2|| + |n_3/n - |\vec{P}_3|| \quad (3)$$

The minimum of the objective function $Z(\vec{P})$ is a minimum for the function $F(\vec{P})$, i.e. both functions reach a minimum for one parameter \vec{P} . We are transforming the expression (3):

$$\min Z(\vec{P}) = |n_1/n - w_1| + |n_2/n - w_2| + |n_3/n - w_3| \quad (4)$$

Expression (4) is a universal tool for adaptive selection of appropriate marking for the traffic flow entering the intersection. That is, from all the markings \vec{P} of Table I we take the marking \vec{P}^* at which the minimum of the function $Z^*(\vec{P}^*)$ is reached. It provides the fastest MV traffic at the crossroads.

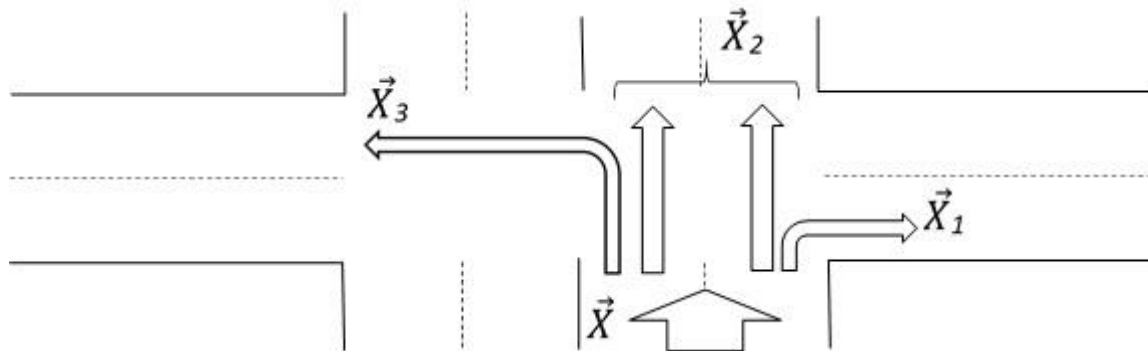



















Fig.4 - Example of a vector traveling bundle representation

IV. CLASSIFICATION OF MARKINGS

We will compose all possible markings for a two-lane entrance to the intersection. The number of all such markings is 17. All the markings obtained can be classified into three groups (Table I):

- Actual.
- Force majeure.
- Irrelevant.

TABLE I
FORMULAS OF TWO-LANE MARKINGS

#	Formula of markup with grouping along the lane	Formula of markup with grouping along the movement direction	Marking numeric characteristics (w_1, w_2, w_3)	Traffic sign
Actual				
1	$(\vec{P}_{11} + \vec{P}_{12}) + (\vec{P}_{22} + \vec{P}_{23})$	$\vec{P}_{11} + (\vec{P}_{12} + \vec{P}_{22}) + \vec{P}_{23}$	$(0.25, 0.5, 0.25)$	
2	$(\vec{P}_{11} + \vec{P}_{12} + \vec{P}_{13}) + \vec{P}_{23}$	$\vec{P}_{11} + \vec{P}_{12} + (\vec{P}_{13} + \vec{P}_{23})$	$(0.16, 0.16, 0.66)$	
3	$\vec{P}_{11} + (\vec{P}_{21} + \vec{P}_{22} + \vec{P}_{23})$	$(\vec{P}_{11} + \vec{P}_{21}) + \vec{P}_{22} + \vec{P}_{23}$	$(0.66, 0.16, 0.16)$	
4	$(\vec{P}_{11} + \vec{P}_{12}) + \vec{P}_{23}$	$\vec{P}_{11} + \vec{P}_{12} + \vec{P}_{23}$	$(0.25, 0.25, 0.5)$	
5	$\vec{P}_{11} + (\vec{P}_{22} + \vec{P}_{23})$	$\vec{P}_{11} + \vec{P}_{22} + \vec{P}_{23}$	$(0.5, 0.25, 0.25)$	
Force majeure				
6	$(\vec{P}_{11} + \vec{P}_{12}) + \vec{P}_{22}$	$\vec{P}_{11} + (\vec{P}_{12} + \vec{P}_{22})$	$(0.25, 0.75, 0)$	
7	$\vec{P}_{12} + (\vec{P}_{22} + \vec{P}_{23})$	$(\vec{P}_{12} + \vec{P}_{22}) + \vec{P}_{23}$	$(0, 0.75, 0.25)$	
8	$(\vec{P}_{12} + \vec{P}_{13}) + \vec{P}_{23}$	$\vec{P}_{12} + (\vec{P}_{13} + \vec{P}_{23})$	$(0, 0.25, 0.75)$	
9	$\vec{P}_{11} + (\vec{P}_{21} + \vec{P}_{22})$	$(\vec{P}_{11} + \vec{P}_{21}) + \vec{P}_{22}$	$(0.75, 0.25, 0)$	
10	$\vec{P}_{11} + (\vec{P}_{21} + \vec{P}_{23})$	$(\vec{P}_{11} + \vec{P}_{21}) + \vec{P}_{23}$	$(0.75, 0, 0.25)$	
11	$(\vec{P}_{11} + \vec{P}_{13}) + \vec{P}_{23}$	$\vec{P}_{11} + (\vec{P}_{13} + \vec{P}_{23})$	$(0.25, 0, 0.75)$	
12	$\vec{P}_{11} + \vec{P}_{23}$	$\vec{P}_{11} + \vec{P}_{23}$	$(0.5, 0, 0.5)$	
13	$\vec{P}_{11} + \vec{P}_{22}$	$\vec{P}_{11} + \vec{P}_{22}$	$(0.5, 0.5, 0)$	
14	$\vec{P}_{12} + \vec{P}_{23}$	$\vec{P}_{12} + \vec{P}_{23}$	$(0, 0.5, 0.5)$	
Irrelevant				
15	$\vec{P}_{13} + \vec{P}_{23}$	$(\vec{P}_{13} + \vec{P}_{23})$	$(0, 0, 1)$	
16	$\vec{P}_{11} + \vec{P}_{21}$	$(\vec{P}_{11} + \vec{P}_{21})$	$(1, 0, 0)$	
17	$\vec{P}_{12} + \vec{P}_{22}$	$(\vec{P}_{12} + \vec{P}_{22})$	$(0, 1, 0)$	

Those markings are actual that use all three outputs from the intersection. These are working markings that are constantly used in the process of controlling the intersection. The frequency of their use is different and depends on the structure of traffic flows crossing the intersection (here the transport stream is considered as the ratio of vehicles turning right, right and left), i.e. the closest marking is set on the electronic scoreboard for the traffic intersecting the intersection. "The closest" means that out of the entire set of markings, only this intersection will leave all cars in the minimum time.

Force majeure markings use only two outputs from the intersection of the three possible. With this marking, one exit from the intersection is closed. This group of markings is not used so often. Only in the case of an accident or during maintenance work to close the passage through the street, which refers to this exit from the intersection.

Out-of-date markup has only one exit from the intersection. This group of markings is practically not used. In Table I, layouts are arranged in decreasing order of use.

Starting from the moving MVs with the characteristics $\vec{X}(n_1, n_2, n_3)$ moving to the intersection at a particular (current) time moment, there is a vector (marking) \vec{P} by searching through Table I to minimize the function $Z(\vec{P})$ based on the expression (4). The search is done by the numerical characteristics of the vector $\vec{P}(w_1, w_2, w_3)$. Thus, an adaptive selection of the closest to the markup flow is carried out, which is constantly replaced by an electronic scoreboard.

V. CONCLUSION

Based on the above analysis of the possibility of using electronic displays with adaptive

algorithm for the integrated intersection management of road traffic in cities, it can be noted, that currently this tool is the most modern and effective. The effectiveness of the proposed system is determined by the possibility of increasing the discipline of drivers and safety on the roads, as well as by economic considerations, manifested in reducing fuel consumption and increasing the resource of cars for drivers and reducing the costs of maintaining standard marking in an acceptable state for public services and the country itself.

REFERENCES

- [1] Yu.A. Vrubel, "Organization of road traffic," Traffic characteristics: a teaching aid for students of specialty 1-44 01 02 "Organization of road traffic," Minsk: BNTU, 2007, 268 p.
- [2] V.M. Markelov, "Intelligent transport systems as a management tool," the State Adviser, 2014, vol. 3, pp. 42-49.
- [3] Yu.A. Kremenets, "Technical means of road regulation, Transport," 1981, pp. 84-116
- [4] S.V. Anfilets and V.N. Shuty, "Adaptive control algorithm based on step-by-step adjustment of traffic light objects along the highway," Reports of BSUIR, Minsk, 2010, pp. 79-86
- [5] V.N. Shuty, "Adaptive management of traffic flows in the city's street-road network, Reports of BSUIR, Publisher BrSTU," 2015, 188 p.
- [6] S.V. Anfilets and V.N. Shuty, "Evaluating The Effectiveness Of The Adaptive Control System In Brest Region," International Congress Of Heavy Vehicles, Road Trains And Urban Transport, 2010, 79-86 pp.
- [7] Road marking. General specifications: STB 1231-2000, Minsk: Belarusian. State. Institute of Standardization and Certification, 2006, p. 82

Application of the Integral Optical Flow for Identification of the Cell Population Motion in the Microscopic Images

Olga Nedzvedz, Sergey Ablameyko, Huafeng Chen

Abstract— In this paper, we propose a method to identify the cell population behavior (directional motion, aggregation and dispersion of cells) for microscopic videos on base of the integral optical flow conception. Integral optical flow is form of accumulation of the basic optical flow. Due to the accumulation, displacement vectors of background become small, while those of foreground keep growing. Based on information extracted from integral optical flow, pixel motions are analyzed statistically at all positions for each frame to obtain quantity of pixels and their comprehensive motion at each position. After that, regional motion indicators are defined and computed to describe motions at region-level. Thresholds for motion intensity, quantity and motion direction of pixels are used together to segment regional motion maps and identify cell population behavior. These properties are useful for further analysis of video scene. Experimental results show that our method can identify this behavior effectively.

Keywords— Integral optical flow, image monitoring, motion analysis, segmentation.

I. INTRODUCTION

Video microscopic technologies have been used for analyzing of cell population behavior and tracking of cells. Using of microscopy provides opportunity for monitoring intact cells. This technique makes the analysis of a cell population completely label-free and gives possibility to study population of thousands of cells in vitro. Automated cell tracking system shall consider that cell populations have varying and inhomogeneous character. Most interesting questions for investigation include dynamical properties of cells and cellular interactions.

Modern video microscopic systems become more intelligent and they allow not only record and visually analyze images but also start to have software tools for tracking, recognition, identification of cells. However, it is necessary to analyze not only behavior of individual cells, but also analyze, trace and predict cells behavior in population, their aggregation and dispersion. Comparing to individual cells, cell population is often more difficult to track.

Identification of cell population behaviors on videos in traditional way is based on the segmentation of objects of interests from the background and track their movements separately [1, 2, 3]. However, for cell populations this is difficult due to their occlusions [4]. Some works choose to study the population as a single entity, but population of cells is a dynamic structure consisting of interacting dynamic elements.

Methods for detection and tracking of dynamic objects can be divided into tracking-based methods that examine trajectories of objects and tracking-free methods based on examination of visual features such as differences in color or brightness changes [3, 5]. Permanent transformation of cell population like change of size, shape and orientation makes difficult using of this method for cells. Detection and tracking of cell population is a more complicated task compared with tracking of separate cells. For cell population it is difficult to determine the boundaries between cells due to their merging.

Methods like thresholding techniques [6, 7], neural networks and probabilistic object models [8, 9] are used for these objects. One of the most promising methods for motion analysis of

Olga Nedzvedz Belarusian State Medical University, Minsk, Belarus (e-mail: olga_nedzved@tut.by).
Sergey Ablameyko Belarusian State University, Minsk, Belarus (e-mail: ablameyko@bsu.by).
Huafeng Chen, Zhejiang Shuren University, Hangzhou, China (e-mail: eric.hf.chen@outlook.com).

dynamic objects based on using an optical flow. Optical flow belongs to region-based methods [11]. This method allows to get the distribution of velocities and directions of points of object from shift of these points between two images. Optical flow is widely used for investigation of different types of motion like translation of moving object relative to the static or dynamic background and another moving objects; rotation of the object relative to the axis. However, case of cell population with unstable shape, internal noises and random changes has a great influence on the basic optical flow in video sequence. As a result, the structure of motion vectors can be unclear.

An important aim of the automated systems is detection of spatio-temporal localization of mitosis events. Every mitosis event is the division of cell into two daughter cells, which is always accompanied by a change in size, shape and brightness of the area around cells. Different mitosis detection methods unclude: mean shift algorithms; multiple-object matching methods based on the frame-by-frame segmentation; tracking algorithms based on determination of blob region's characteristics; methods of detection of mitosis based on brightness change [11, 13]. Another classification of approaches to detection of mitosis divides all methods on temporal and spatio-temporal [12]. According to this classification, temporal methods detect moment of time, when one cell divides into two cells. Spatio-temporal methods detect as size, shape, velocity and brightness of cells change.

In this study, we propose a new method to identify cell population behavior based on integral optical flow. Unlike basic optical flow [1, 3, 8] by which foreground is hard to pull apart from background due to random motion of the background, integral optical flow enhance foreground motion and restrict background motion. Therefore, foreground is easy to be obtained. Our method considers three factors to identify motion: motion intensity, quantity and motion direction of pixels moving toward and moving away from certain regions.

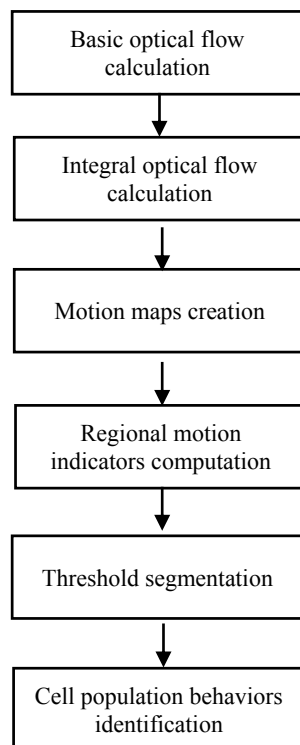


Fig. 1 General scheme of population behaviors identification

In this study, we determine basic structures for three types of motion: directional motion, aggregation of cells and dispersion of cell population. Based on integral optical flow, we

describe motion on the pixel level and at region level. At last, we use threshold segmentation to identify above-mentioned cell conglomerate behaviors. We apply our method on microscopic videos and get good results. General scheme of our method is shown in Fig. 1.

II. INTEGRAL OPTICAL FLOW

Optical flow describes the movement of pixels between two consecutive frames. The idea of optical flow is based on two assumptions: (1) the pixel intensities of an object do not change between consecutive frames; (2) neighboring pixels have similar motion.

A video frame is a digital image and there is an imaginary underlying grid below it. Each coordinate point of the grid is called a position. Positions are stationary, while pixels may change their positions from frame to frame. Optical flow provides a way letting one study pixel motion at each position at different time.

Basic optical flow only records displacement vector of pixels between two consecutive frames. Taking into account the very short time interval between them, it is hard to distinguish foreground from background due to motion of background. Generally background moves randomly, e.g. back and forth or rotationally. In short time this kind of character doesn't show, but after a long enough time, it will reveal itself thus help identify foreground.

Integral optical flow is an intuitive idea that accumulate optical flow for several consecutive frames. In result of this accumulation, displacement vectors of background become small, while those of foreground keep growing [9].

For description convenience, we use I_t to denote t -th frame of video I , $I_t(p)$ to denote pixel at position p in I_t . Let OF_t denote basic optical flow of I_t . It is a vector field with each vector $OF_t(p)$ represents displacement vector of pixel $I_t(p)$. Assume, $OF_t(p) = \vec{d}$ we can easily determine the position in I_{t+1} at which pixel $I_t(p)$ moves to, which will be $p + \vec{d}$.

Consider optical flows for several consecutive frames have been computed, we can obtain integral optical flow for the first frame of those. Let IOF_t^{itv} denote integral optical flow of I_t , where itv is the frame interval parameter used to compute integral optical flow. IOF_t^{itv} is also a vector field which records accumulated displacement information in time period of itv frames for all pixels in I_t .

For any pixel $I_t(p)$, its integral optical flow $IOF_t^{itv}(p)$ can be determined as follow:

$$IOF_t^{itv}(p) = \sum_{i=0}^{itv} OF_{t+1}(p_{t+1})$$

where p_{t+1} is the position in I_{t+1} of pixel $I_t(p)$. Note that x-component and y-component of p_{t+1} should be rounded to the nearest integer, as pixels are at positions with integer coordinates.

Integral optical flow allows to identify foreground or mobile regions by threshold segmentation. Any pixel with a big enough magnitude of its displacement vector will be considered as a foreground pixel.

III. IDENTIFICATION OF CELL POPULATION BEHAVIOR

A. Types Characteristics of Cell Population Motion

We consider three types of cell population motions: directed motion of the cell population, aggregation of cells and dispersion of population.

Directed motion of cells is identified in the case when cells or a cell population moves in the same direction. There are main features of directed motion: some objects move simultaneously from one area of the image to another one; the speed of objects exceeds the speed of chaotic background movement; the direction of movement of cells is the same (Fig. 2a).

Cells aggregation is identified when cells move into a certain region from different directions, this motion can be symmetrical relative to the directions of motion or not. There are main features of aggregation: some objects moves into a certain region from elsewhere; the speed of motion is greater than the speed of chaotic background movement; there are at least two moving directions and they are more or less symmetrical (Fig. 2b).

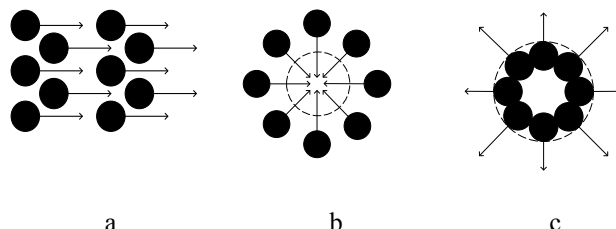


Fig 2 Diagram of different types of cells motion: (a) directed motion; (b) aggregation; (c) dispersion. Black solid circles represent cells, arrows show motion directions

Cells dispersion is identified when cells move out of a certain region to different directions like it shown in Fig. 2c. As example, it can be growing of cell population or cell scattering under the influence of some negative factors. Cells dispersion happens when cells move out of a certain region to elsewhere, the speed of motion is greater than the speed of chaotic background movement and their motion is more or less symmetrical.

In general situation it is possible to describe the motion of set of objects with help of a concept of position or a concept of a region, which includes a set of adjacent positions. Basic optical flow and integral optical flow record basic information of motions: starting positions and ending positions of pixels, it allows to determine comprehensive moving directions of pixels.

Motion direction indicates a destination where cells move. In order to determine cells motion direction in a region, we can simply divide $[0, 2\pi)$ into several intervals with equal length and count for each interval number of pixels whose motion direction is in that interval. Interval with most pixels shows main motion direction of cell population.

Motion speed of pixel $I_t(p)$ in time period from I_t to I_{t+1} can be calculated as follow:

$$s_t^{itv}(p) = \frac{|IOF_t^{itv}(p)|}{itv}$$

B. Motion Analysis based on the Pixel-level

Integral optical flow allows to determine motion of a pixel in given time period. In Fig. 3a pixel in I_t with original position p_0 moves to position p_3 after four frames in I_{t+4} . Position sequence (p_0, p_3) is simple motion path of the pixel in time period from frame I_t to frame I_{t+4} . In Fig. 3b position sequence (p_0, p_3) is interpolative motion path of the pixel in the same time period. We use an interpolative motion path along with a simple motion path for calculation of integral optical flow.

$\vec{D}_1, \vec{D}_2, \vec{D}_3, \vec{D}_4$ are displacement vectors of the pixel extracted from basic optical flows $OF_t, OF_{t+1}, OF_{t+2}, OF_{t+3}$, respectively. $\vec{D}'_1, \vec{D}'_2, \vec{D}'_3, \vec{D}'_4$ in Fig. 3b are their integer versions extracted from integral optical flow IOF_t^4 :

$$\vec{D} = \vec{D}'_1 + \vec{D}'_2 + \vec{D}'_3 + \vec{D}'_4$$

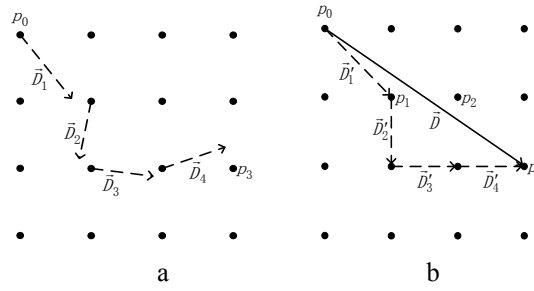


Fig 3 Pixel motion path: (a) simple motion path; (b) interpolative motion path

Consider a vector $\overrightarrow{p_0 p_{n-1}}$ represents an integral optical flow for the effective pixel motion path starting from p_0 and ending at p_{n-1} where $n \geq 2$. In order to compute contributions of pixel quantity and comprehensive motion of this path for every position in it, we use the normalized vector of integral optical flow, because it carries information of both pixel number and motion direction. The normalized vector of $\overrightarrow{p_0 p_{n-1}}$ can be computed as follow:

$$\overrightarrow{v_{norm}} = \frac{\overrightarrow{p_0 p_{n-1}}}{|\overrightarrow{p_0 p_{n-1}}|}$$

where $\overrightarrow{v_{norm}}$ is the normalized vector, $|\overrightarrow{p_0 p_{n-1}}|$ is the magnitude of $\overrightarrow{p_0 p_{n-1}}$.

Vector $\overrightarrow{v_{norm}} = 1$ indicates pixel number for a single effective pixel motion path, four values can be computed for any position p_i ($0 \leq i < n$) in the path:

$$\begin{aligned} s_{in} &= w_{in} \cdot |\overrightarrow{v_{norm}}|; \\ s_{out} &= w_{out} \cdot |\overrightarrow{v_{norm}}|; \\ \overrightarrow{v_{in}} &= w_{in} \cdot \overrightarrow{v_{norm}}; \\ \overrightarrow{v_{out}} &= w_{out} \cdot \overrightarrow{v_{norm}}; \end{aligned}$$

where w_{in} , w_{out} are weighting coefficients for determining percentage of incoming and outgoing pixels at the position and $w_{in} + w_{out} = 1$.

For statistical motion analysis, only pixels, which actually move, should be considered. Thus for each position, only pixel motion path whose starting position is different from its ending position should be taken into account.

Description of cell motion and determination of the type of motion can be performed with help of motion maps. In this paper, several types of motion maps are proposed. Suppose that in-pixel quantity map (*IQ-map*) is a map with a scalar value at each position indicating number of pixels moving toward the corresponding position. Out-pixel quantity map (*OQ-map*) is a map with a scalar value at each position indicating number of pixels moving away from the corresponding position. In-pixel comprehensive motion map (*ICM-map*) is a map with a vector at each position indicating comprehensive motion of pixels moving toward the corresponding position. Out-pixel comprehensive motion map (*OCM-map*) is a map with a vector at each position indicating comprehensive motion of pixels moving away from the corresponding position.

IQ-map, *OQ-map*, *ICM-map* and *OCM-map* of I_t can be determined as follows to formulas:

$$\begin{aligned} IQ_t(p) &= \sum_{a \in S_t(p)} s_{in}(a, p) \\ OQ_t(p) &= \sum_{a \in S_t(p)} s_{out}(a, p) \\ ICM_t(p) &= \sum_{a \in S_t(p)} \overrightarrow{v_{in}}(a, p) \end{aligned}$$

$$OCM_t(p) = \sum_{a \in S_t(p)} \overrightarrow{v_{out}}(a, p)$$

where p is value of position on appropriate map, $S_t(p)$ is set of motion paths (simple or interpolative depending on which type of motion path is used for motion analysis).

According to created motion maps for time of I_t , important characteristics of pixel motions at that time will be determined: positions with bigger value on *IQ-map* are positions to which more pixels move; positions with bigger value on *OQ-map* are positions from which more pixels leave; positions with smaller vector magnitude on *ICM-map* are positions to which pixels move in more symmetrical directions; positions with smaller vector magnitudes on *OCM-map* are positions from which pixels leave in more symmetrical directions. Based on this, one can conclude that positions with big values on *IQ-map* and small vector magnitudes on *ICM-map* are positions at which pixels tend to aggregate; positions with big values on *OQ-map* and small vector magnitudes on *OCM-map* are positions at which pixels tend to disperse.

C. Motion Analysis based on Region-level

Description of the cell population motion one can perform at the region-level based on the study of the displacement of regions of interest. Motion intensity is a major factor for detecting different events in cell population. Any region with a high enough regional motion intensity is considered an intensive motion region. It is appropriate to use average displacement of pixels to represent motion intensity in a region.

Regional motion intensity is an average of displacement vector magnitudes extracted from integral optical flow for pixels in a certain region. It can be described according to formula:

$$MI_t(r) = \frac{1}{n} \sum_{p \in r} |IOF_t^{itv}(p)|$$

where $MI_t(r)$ is the regional motion intensity of region r at the time of I_t , n is position number in r , p is one position in r , $IOF_t^{itv}(p)$ is the displacement vector of $I_t(p)$ extracted from integral optical flow.

For any region regional in-pixel relative quantity $IRQ_t(r)$ of certain region r at the time of I_t can be determined as follow:

$$IRQ_t(r) = \frac{1}{n} \sum_{p \in r} IQ_t(p),$$

where r is certain region, $IQ_t(p)$ is value at position p on *IQ-map* at the time of I_t .

Regional out-pixel relative quantity $ORQ_t(r)$, which is an average of values on *OQ-map* at positions in a certain region, can be determined by formula:

$$ORQ_t(r) = \frac{1}{n} \sum_{p \in r} OQ_t(p)$$

where $OQ_t(p)$ is value at position p on *OQ-map* at the time of I_t .

By comparing $IRQ_t(r)$ with $ORQ_t(r)$, one can know whether more pixels move toward a certain region than pixels move away from it or vice versa. Identification of cells aggregation or cells dispersion can be performed by calculation regional in/out indicator IOI_t :

$$IOI_t = \frac{IRQ_t(r)}{ORQ_t(r)}$$

In this formula $IOI_t(r) > 1$ means more pixels move into r , while $IOI_t(r) < 1$ means more pixels move out of r .

There are two additional features to describe the nature of the motion of a cell population. Regional in-pixel symmetry $IS_t(r)$ is the ratio of regional in-pixel relative quantity of a certain region to magnitude of average vector of that region on $ICM-map$:

$$IS_t(r) = \frac{IRQ_t(r)}{\left| \frac{1}{n} \sum_{p \in r} ICM_t(p) \right|}$$

where n is position number in r , $IRQ_t(r)$, $ORQ_t(r)$ are values for region r on $IRQ-map$ and $ORQ-map$ at the time of I_t , respectively, $ICM_t(p)$, $OCM_t(p)$ are values at position p on $ICM-map$ and $OCM-map$ at the time of I_t , respectively.

Regional out-pixel symmetry $OS_t(r)$ can be determined by the formula respectively:

$$OS_t(r) = \frac{ORQ_t(r)}{\left| \frac{1}{n} \sum_{p \in r} OCM_t(p) \right|}$$

The bigger $IS_t(r)$ or $OS_t(r)$ is, the more symmetrically corresponding pixels move.

IV. IDENTIFICATION CELL POPULATION BEHAVIOR USING MOTION MAPS

The identification of the behavior of cell population in the region r can be performed on base of the following characteristics: regional motion intensity $MI_t(r)$, regional out-pixel relative quantity $ORQ_t(r)$ and regional out-pixel symmetry $OS_t(r)$. Lower limits of region size and regional motion intensity should be properly determined for this identification according to specific properties of cells.

Directed motion occurs if $MI_t(r)$, $ORQ_t(r)$ and $OS_t(r)$ equal to the thresholds or exceed them. Thereby, the directional motion is going to happen in region r at the time of I_t if the following conditions are met:

$$MI_t(r) \geq t_{11}; ORQ_t(r) \geq t_{12}; (3) OS_t(r) \geq t_{13}.$$

where t_{11} is the threshold for $MI-map$, t_{12} is the threshold for $ORQ-map$, t_{13} is the threshold for $OS-map$.

The identification of cells aggregation can be performed on base of the following characteristics: regional motion intensity $MI_t(r)$, regional in-pixel relative quantity $IRQ_t(r)$, regional in/out indicator $IOI_t(r)$, regional in-pixel symmetry $IS_t(r)$. Aggregation occurs if $MI_t(r)$, $IRQ_t(r)$, $IOI_t(r)$, $IS_t(r)$ meet thresholds:

$$MI_t(r) \geq t_{21}; IRQ_t(r) \geq t_{22}; IOI_t(r) \geq t_{23}; IS_t(r) \geq t_{24},$$

where t_{21} is the threshold for $MI-map$, t_{22} is the threshold for $IRQ-map$, t_{23} is the threshold for $IOI-map$, t_{24} is the threshold for $IS-map$.

Cells dispersion is identified in region r if $MI_t(r)$, $ORQ_t(r)$, $IOI_t(r)$ and $OS_t(r)$ meet thresholds:

$$MI_t(r) \geq t_{31}; ORQ_t(r) \geq t_{32}; IOI_t(r) \geq t_{33}; OS_t(r) \geq t_{34},$$

where t_{31} is the threshold for $MI-map$, t_{32} is the threshold for $ORQ-map$, t_{33} is the threshold for $IOI-map$, t_{34} is the threshold for $OS-map$.

Threshold for motion intensity can be determined with help of the frame interval parameter. With bigger frame interval parameter, more intensive motion field will be obtained, thus threshold for motion intensity should be bigger for the same situation monitoring application. It can be determine as:

$$t_s = \alpha \cdot itv$$

where t_s is threshold for motion intensity, itv is frame interval parameter, α is a constant for the specific task.

Region size is another parameter that affects determination of threshold. If bigger region is used, threshold for pixel quantity should be smaller:

$$t_q = \frac{\beta}{s}$$

where t_q is threshold for pixel quantity, s is area of region, β is a constant for the specific task.

V. STAGES OF THE CELL CYCLE DETECTION

The method based on the integral optical flow allows detecting of spatio-temporal localization of cell cycle stages. We consider four main cell cycle stages: normal stage, mitosis, apoptosis and intermediate stage. The information about cell stage is significantly important for the monitoring of cell population evolution. Every change in the stage of cell cycle is accompanied by a change in size, shape and brightness of the area around cells. Processes of mitosis or apoptosis are going from intermediate stage. Vectors of integral optical flow going from center of any event have structure like a star. During the intermediate stage of evolution cell is growing and vectors of optical flow have a star structure and after that cell can be destroyed (in stage of apoptosis) or divided into two new cells (in stage of mitosis). In case of apoptosis, vectors of optical flow have various random directions. In case of mitosis, they have two opposite directions.

Determination of cell cycle stages can be performed with help of additional motion map called regional out-pixel comprehensive motion map (*ROCM-map*), can also be created to show how symmetrically pixels move away from a region. Let $ROCM_t$ denote ROCM map, then

$$ROCM_t(c) = ROCM_t(r) = \frac{1}{n} \sum_{p \in r} OCM_t(p),$$

where n is position number in r , c is the center of cell, which coincides with center of the region.

In general, the bigger $ROCM_t(c)$ is, the more symmetrically corresponding pixels move. For mitosis stage symmetry on the *ROCM-map* is larger than the threshold value, for apoptosis stage it is very low.

VI. CONCLUSION

We have presented a method for cell population behaviors identification. Our method mainly consists of the following steps: integral optical flow computation, position-level motion analysis, region-level motion analysis and threshold segmentation. The accumulative effect of integral optical flow allows to obtain steady motion regions, which are usually regions of interest. Pixel motion intensity, quantity and motion direction are together used to describe motions and identify cell population behaviors. The effectiveness of our method has been demonstrated and confirmed by our experimental results. Due to the combination of simple solutions, this algorithm can be easily realized on base of many computer systems of image analysis.

Although we can get good results for jittering videos through changing thresholds in some cases, but that is not universal, thus our method should be applied to stable videos generally. In different applications thresholds for motion intensity, quantity and motion direction of pixels should be set depending on the scene, camera setup and specific purposes.

ACKNOWLEDGMENT

This work is supported by Belarusian Republican Foundation for Foundational Research (F16R-180), Zhejiang Provincial Natural Science Foundation of China (LQ12D01004), the National High-end Foreign Experts Program (GDW20163300034), and Public Welfare Technology Applied Research Program of Zhejiang Province (2015C33074, 2015C33083).

REFERENCES

- [1] T. Kanade, Z. Yin, R. Bise, S. Huh, S. Eom, M. F. Sandbothe, M. Chen, "Cell image analysis: Algorithms, system and applications", in: IEEE Workshop on Applications of Computer Vision, pp. 374 – 381, 2011.
- [2] L. Kang, E. Miller, M. Chen, T. Kanade, L. Weiss, P. Campbell, "Cell population tracking and lineage construction with spatiotemporal context", *Medical Image Analysis*, № 12, pp. 546–566, 2008.
- [3] M. Tao, J. Bai, P.Kohli, "SimpleFlow: A Non- iterative, Sublinear Optical Flow Algorithm", *Computer Graphics Forum*. Blackwell Publishing Ltd, 31(2pt1), p. 345-353, 2012.
- [4] G. Yuan, J. Zhang, Y. Han, H. Zhou, D. Xu, "A multiple objects tracking method based on a combination of camshift and object trajectory tracking", *ICSI 2015. LNCS*, vol. 9142, pp. 155–163.
- [5] O. Elharrouss, D. Moujahid, S.E. Elkaitouni, H. Tairi: "Moving objects detection based on thresholding operations for video surveillance systems". In: *IEEE/ACS 12th International Conference of Computer Systems and Applications*, Marrakech, pp. 1–5, 2015.
- [6] C. Qiu, Z. Zhang, L. Huanzhang, H. Luo, "A survey of motion-based multitarget tracking methods", *Prog. Electromagnet. Res. B* 62, pp. 195–223, 2015.
- [7] P. Vijayakumar, A.V. Senthilkumar: "Threshold based filtering technique for efficient moving object detection and tracking in video surveillance. *Int. J. Res. Eng. Technol.* 5(2), pp. 303–310, 2016.
- [8] J. Yang, H. Li: "Dense, accurate optical flow estimation with piecewise parametric model", in: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1019–1027, 2015.
- [9] C. Chen, S. Ye, H. Chen, O. Nedzvedz, S. Ablameyko, Integral optical flow and its applications for dynamic object monitoring in video, *Journal of Applied Spectroscopy*, 84 (1), 120-128 (2017)
- [10] S. Huh, S. Eom, et al. Mitosis detection of hematopoietic stem cell populations in time-lapse phase-contrast microscopy images. *Processing of the 9th IEEE International Symposium on Biomedical Imaging (ISBI)*, 2012, pp.390-393.
- [11] A-A Liu, K. Li, and T. Kanade, Mitosis Sequence Detection using Hidden Conditional Random Fields, in *Proc. IEEE International Symposium on Biomedical Imaging*, 2010. pp. 580-583.
- [12] M. Maska, O. Danek, et al. Segmentation and Shape Tracking of Whole Fluorescent Cells Based on the Chan- vese Model. *IEEE Transactions on Medical Imaging* vol .32, no. 6, pp. 995 - 1006.
- [13] R. Nathiya and G. Sivaradje. An effective mitosis recognition and segmentation tool for stem cell exploration. *ARPJ Journal of Engineering and Applied Sciences*, vol. 10, no. 6, 2015, pp.2691-2696

Mobility and Public Transport Service: from Deurbanization to Decentralization

Valeriy Marunich, Oleksandr Melnichenko, Vitalii Kharuta

Abstract—The aim of the study is to demonstrate the methodology for improving commuter traffic in the context of the decentralization of transport infrastructure concept and on this basis to increase mobility and transport services for the rural population with generally accepted European social standards and guarantees.

Keywords—transport system, transport infrastructure, deurbanization, decentralization, rural population

I. Introduction

General problem definition and its relationship with important scientific and practical tasks.

The uniqueness of the Ukrainian communication and mobile problems is that no other country in the world has transferred transport system to a new social level of quality yet. This work should be preceded by a scientific program approach and a clear legislative framework. As for passenger transportations, our national experts have been doing some work on improving commuter bus transportation. But they solved only one or some problems without systematic approach [1]. So, the structure of demand for suburban transportation has not been investigated, because the problems of the particular region were solved and a wide implementation was not summarized. There was an attempt to analyze the results of suburban bus services based on their technical standards and operational parameters [2], and not dynamic but static traffic system was studied without considering patterns of passenger traffic flow.

Note that studying the impact of individual and public transport development on the economic characteristics of the transport system [3], foreign experts concluded that in suburban and rural areas a reduction of transport service means a reduction of buses or its complete termination, resulting in the need of walking to the nearest functioning route. As a result, most passengers prefer their own transport, which otherwise they would not buy. Thus, the relative cost of travel tends to increase. For example, in the UK there is a deurbanization tendency (urbanization - (from Lat. Urbanus – city) – the growing importance of cities for the development of society, which is accompanied with growth and development of urban areas, increasing the proportion of the urban population, the spread of urban lifestyle in a particular region, country, world), leading to the distribution of the working population habitats in a large area with low concentration. This may explain the fact that in recent decades the number of personal cars has increased almost 8 times, and the number of trips made by suburban public transport compared to the 60s decreased to 50%. This trend has meant that in England the family accounts for more than one car.

Nowadays functioning transport system of Ukraine where the level of individual transport density is much lower, taking into account imperfections of automotive manufacturing and transport complex [4], the development of public transport remains prioritized. After all, if the EU countries try to improve mobility with the help of wide application of deurbanization trends, in Ukraine the concept of power decentralization has been launched and takes practical shape, which is regarded as the transfer of significant powers and budgets of state agencies to local governments so that the bodies closer to the people, where such powers can be implemented most effectively, could have as much authority as possible. [5] As for the increased mobility of Ukrainian population, scientific and practical support for the implementation of the of transport infrastructure is urgent.

V. Marunich National Transport University, Kiev, Ukraine (e-mail: marunic@univ.kiev.ua)

O. Melnichenko National Transport University, Kiev, Ukraine (e-mail: melnichenko@ntu.edu.ua).

V. Kharuta National Transport University, Kiev, Ukraine (e-mail: kharuta_vitaliy@ukr.net).

Analysis of recent research and publications. Within the framework of the European Program, Transport White Paper 2011 issues of increasing mobility and transport services for the population occupy the top position.

The European Commission has adopted a plan of 40 specific initiatives over the next decade to build a competitive transport system that will increase mobility, remove major barriers in key areas, dramatically change Europe dependence on imported oil and will reduce carbon emissions in transport by 60% till 2050 [6].

These issues are important for Ukraine and are now at the stage of project and program development. In particular, the Supreme Council of Ukraine regulation of 08.31.2015 Number 656-VIII "About preliminary approval of the bill draft on amendments to the Constitution of Ukraine concerning the decentralization of power" in the Policy brief "Transport Policy of Ukraine and its approximation to EU standards" and" regarding the priority measures to overcome the crisis in the transport system of Ukraine, in other laws and regulations important theoretical and practical aspects on this issue need proper software and organizational and financial support [7-8].

Highlighting unsolved aspects of a general problem, which the article is devoted to. The absence of single system approach and methodology for improving bus transport in suburban mobility to enhance transport services for the rural population in the context of decentralization of transport infrastructure concept.

II. Simulated Annealing

The main part with full justification of scientific results. Scientists of National Transport University developed a system of methods for improving bus transport of commuter traffic. Its main elements are the modern information support, modeling and evaluation of existing transport infrastructure, design of transportation system, operational management of transport process and its economic security. It is proved that the principles of the system suggest that existence of the road network and its combination are defined with areas and quantities of passenger traffic flow when it is regulated by the intensity of bus traffic and the fees for transportation.

For example, the analysis of commuter route system in Zaporizhya region revealed that settlements are usually directly connected with district and regional centers, that is radial nature of transport links is distinctly evident. Determination of tariffs depending on the passenger traffic distance doesn't make it possible to use optimization models of urban transport system.

Let us consider fragmented options for the design of route system (Fig. 1). Thus, when considering the standard version (Fig. 1 a) there is possible connection of settlements O, K, P, C with ring route O-K-P-C-O or with radial routes O-K-P, O-R-O, O-C-O and some other combinations. In the fragment of Fig. 1. b the connection between settlements O and P is possible in terms of separate route O-K-P-O or the route with a stop at the point P, so the route is O-K-P-K-C-O.

Thus, justifying route system, the problem of selecting appropriate options of route formation arises. For this purpose economic and mathematical model of rational number of route determination on the minimum social costs criterion was proposed. It should, first of all, evaluate the time spent by passengers for waiting and travelling by buses. For convenience some options of the routes are given (Fig. 1) with the passenger traffic flow pattern. In this case, public service can be organized in four versions: the first route involves O-K, O-P, O-C; the second one is O-K, O-C; the third one is O-P, O-C; the fourth one is O-C. According to the distribution of passenger traffic flows based on the direction, technical and operational parameters of vehicles, it is necessary to identify the most appropriate version of the route system. For this the multicriterion problem is solved with the need to minimize the number of operated vehicles, and the time spent by passengers on waiting and travelling by bus.

Analysis of the route system from a social point of view, based on hours spent by passengers on waiting and travelling, is carried out using the following ratio:

$$\begin{aligned}
 & \left(\sum_{N^{//}} \Pi_i^{//} t_{iwait.}^{//} + \sum_{N^{//}} \frac{\ln_i}{V_{ci}} \right), \\
 & \left(\sum_{N_i} \Pi_i t_{iwait.} + \sum_{N_i} \frac{\ln_i \Pi_i}{V_{ci}} \right).
 \end{aligned}
 \tag{1}$$

where: Π_i is the passenger traffic flow on i route; $t_{iwait.}$ is the time spent by passengers waiting on i route, hours; \ln_i is the average distance of travel on i route, kilometers; V_{ci} is the speed of communication on i route, kilometers per hour; $//$ are the indices appropriate for route systems that are analyzed.

It is possible to analyze the dependence of social costs on the number of commuter routes, using economic and mathematical model that takes into account running costs and level of transport services.

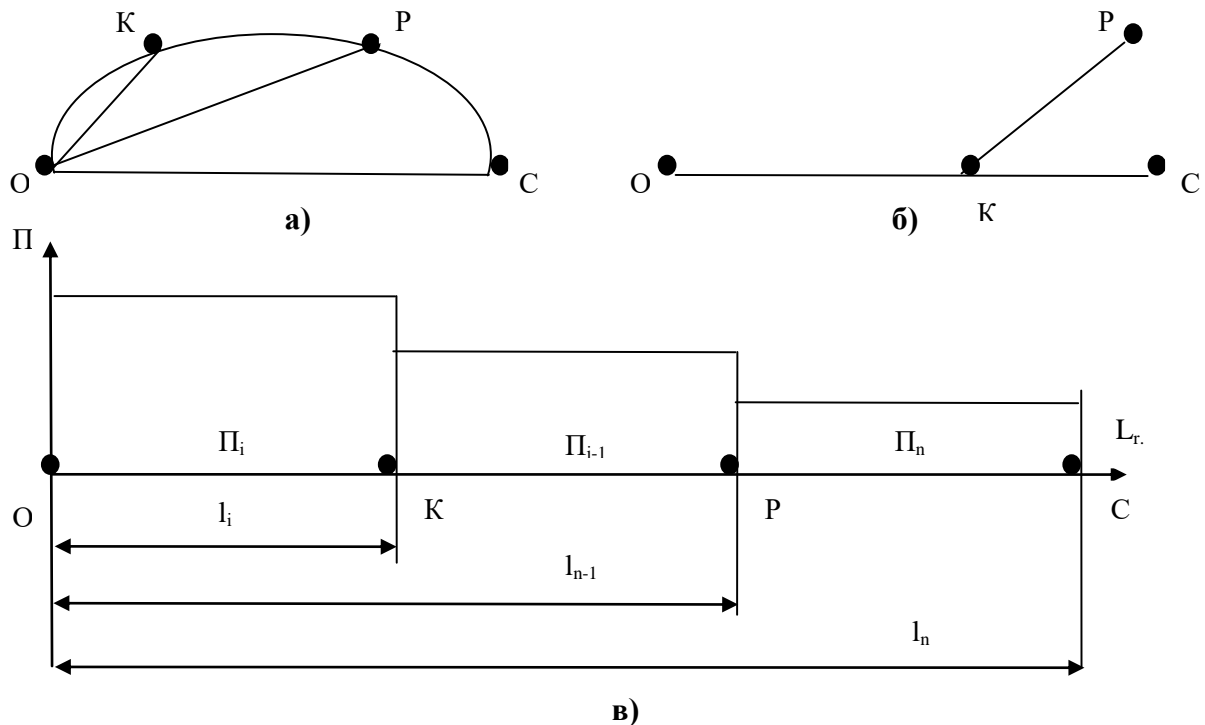


Fig. 1 – the graph of transport network and the diagram of passenger traffic flow on commuter routes: a) b) – the typical options; c) – generalized chart of passenger traffic flow distribution

To build economic and mathematical model we used baseline data and calculation standards. First of all, they include information about the size and distribution of passenger traffic flows, structure and characteristics of the route system, the speed of communication, characteristics of station and road network, type of vehicle and so on. Economic and mathematical model determines the number of routes depending on the estimated number of routes. Logistics construction of such model allows for a wide range of routes in definite directions with different transportation capacity and with any correspondence of passenger traffic flow depending on the social costs, part of which is directly proportional, and some inversely proportional to the number of routes. This makes it possible to define a minimum social costs that meet the optimal number of routes. Economic and mathematical model is as follows:

$$E_{\mathcal{M}} = \frac{C_1}{N} + \frac{C_2}{N} + \frac{C_3}{N} + C_4 N + C_5 N + C_6 N + C_7 N,
 \tag{2}$$

where: C_1 is expenses for fueling and oiling, wearing and tire repair, hryvnas; C_2 is running costs, expenses for maintenance and repair of vehicles, hryvnas; C_3 is expenses for drivers' and conductors' salary and wage, hryvnas; C_4 is the costs of bus stops including forward and reverse directions, hryvnas; C_5 is the time spent by passengers waiting for buses and going by them, hryvnas; C_6 is the costs for reconstruction and maintenance of roads and bus stations, hryvnas; C_7 is the costs of travel for a fee, hryvnas; N is the number of routes, numbers.

Solving the problem of extreme searching for a minimum cost function of the route number is reduced to differentiate the above equation for N , that means that:

$$\frac{E_M}{N_M} = \frac{C_1}{N^2} + \frac{C_2}{N^2} + \frac{C_3}{N^2} + C_4 + C_5 + C_6 + C_7, \quad (3)$$

or,

$$C_1 + C_2 + C_3 = (C_4 + C_5 + C_6 + C_7)N^2, \quad (4)$$

so,

$$N = \sqrt{\frac{C_1 + C_2 + C_3}{C_4 + C_5 + C_6 + C_7}}, \text{ routes.} \quad (5)$$

The following economic and mathematical model is implemented under conditions of dynamic information change on demand for transport. This approach is sufficient to account for the features of the complex object, which is the suburban bus transport system.

The model built in the information complex of technologies and ways of bus transportation management in suburban traffic.

The investigation of relationship between the commuter routes number and transportation process (Fig. 2). found that with increasing capacity of passenger traffic flow and length of routes their number increases. And with the increasing cost of one passenger hour the required number of routes is reduced.

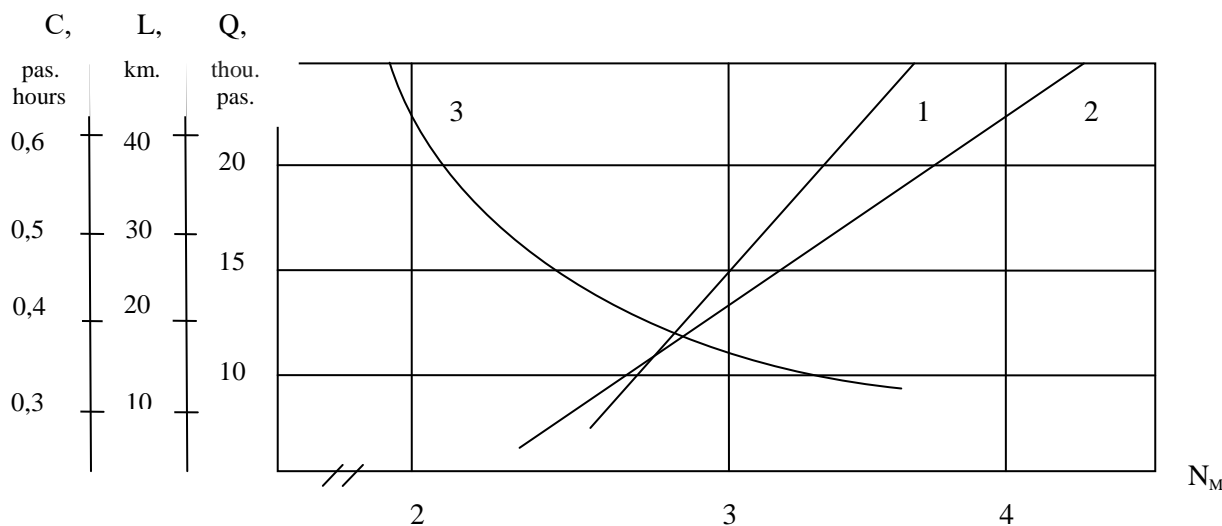


Fig. 2 — the graph shows number of routes 1 — passenger traffic flow capacity; 2 — the length of routes; 3 — passenger hours

III. Low Shop Problematic

The research of implementation of improving commuter traffic methods was carried out on the example of Zaporizhzhya region. Almost for the first time a comprehensive study of the passenger correspondence matrix method use took place in 16 districts of the region during four specific days. 147 inner-suburban and 87 intercity routes were examined. The computer processing required about 560 000 Kb of RAM. Background information was the basis for the analysis of transport services. For example, in the Melitopol district of 70 settlements (1 town, 1 urban village and 68 villages) with the number of inhabitants of about 240 thousand 23 intra-suburban routes operated, covering 91% of the populated areas in the district. Only 3 thousand people living in 6 settlements are not covered with transport. According to these characteristics the level of public service is satisfactory, but the shortage of fuel resources and the availability of vehicles does not allow the system to operate in the required mode. In this respect, because of the available resources, some recommendations were elaborated to involve optimization of suburban routes from 23 to 18 and the number of travels from 145 to 132 per day. This became the basis of the fuel and financial balance, which determined that for the optimal organization of transport it is necessary to provide monthly fuel provision of 18.1 thousand liters of petrol and 17.2 thousand liters of diesel.

Computerization and automation of solving these problems of improving commuter bus transportation reduced the complexity and performance, and the dynamic simulation model of transport system was received. Thus, the given research shows the urgency of commuter traffic improvement and requires further scientific methods of solving them.

IV. Methods

Moreover, studies of the optimal route construction for the commuter routes using the theory of graphs, found that one and the same set of transport network connections can be served by different number of buses, that makes us focus on the passenger time distribution, which means the formation of traffic schedule using Gantt charts and Bayesovskiy methods, etc. (Table. 1).

TABLE 1

The graph and matrix of transport links for route system in time (including range of motion)

Transport network fragment (Gantt chart)	Variations of transport links	Graph and matrix of transport links	Graph of transport network based on adjacency matrix
		$G = \begin{bmatrix} g_{01}, g_{13}, g_{34} \end{bmatrix}$	 $P = \begin{bmatrix} 0134 \\ 0111 \\ 1011 \\ 1101 \\ 1110 \end{bmatrix} \begin{matrix} 0 \\ 1 \\ 3 \\ 4 \end{matrix}$
		$G = \begin{bmatrix} g_{02} \end{bmatrix}$	 $P = \begin{bmatrix} 02 \\ 01 \\ 10 \end{bmatrix} \begin{matrix} 0 \\ 2 \end{matrix}$
		$G = \begin{bmatrix} g_{01}, g_{13} \end{bmatrix}$	 $P = \begin{bmatrix} 013 \\ 011 \\ 101 \\ 110 \end{bmatrix} \begin{matrix} 0 \\ 1 \\ 3 \end{matrix}$
		$G = \begin{bmatrix} g_{01} \end{bmatrix}$	 $P = \begin{bmatrix} 01 \\ 01 \\ 10 \end{bmatrix} \begin{matrix} 0 \\ 1 \end{matrix}$
		$G = \begin{bmatrix} g_{02} \end{bmatrix}$	 $P = \begin{bmatrix} 02 \\ 01 \\ 10 \end{bmatrix} \begin{matrix} 0 \\ 2 \end{matrix}$
		$G = \begin{bmatrix} g_{03} \end{bmatrix}$	 $P = \begin{bmatrix} 03 \\ 01 \\ 10 \end{bmatrix} \begin{matrix} 0 \\ 3 \end{matrix}$

Statement: one and the same known number of travels may be served by different number of buses

V. Results

The conclusions of this study and further research conducted in this direction.

Implementation of the proposed developments in the Zaporizhya region allowed: to set up transportation according to demand and considering resources provision, thus the mobility of the rural population was significantly improved; to streamline the route system, making it possible to reduce the number of routes from 221 to 150 while reducing daily volume of buses; to increase speed of public transport connection by 10-15%; to decrease fuel consumption of passenger transportation by 25% and to reduce emissions of harmful substances into the environment by 15%.

References

- [1] E. M. Lysenkov, "Passenger transportation in a suburban bus service: Overview. (Periodical style)," *Inform. CBNTI*. – M., 1986. – Issue. 2. – P. 54.
- [2] V. F. Shtanov, "Increasing efficiency of passenger transportation in the Ukrainian SSR: Review (Periodical style)," *Inform. CBNTI*. – M., 1988. – Issue. 6. – P. 11-13.
- [3] B. Simpson, *Planning of urban development and public transport in the UK, France and Germany* (Book style). Moscow: Transport, 2000. – 96 p.
- [4] A. M. Redzyuk, "The concept of automotive and motor systems of Ukraine: (Periodical style)," *Avtoshlyahovyk Ukraine*. – K., – 1992. – №3. – P. 2-5.
- [5] About the state social standards and the state social guarantees: Law of Ukraine dated October 5, 2000 No. 2017-III (Electronic resource). – Access mode: <http://zakon2.rada.gov.ua/laws/show/2017-14>.
- [6] European strategies "Transport White Paper 2011".
- [7] T. Syriychik, A. Furgalsky, Ed. M. Svenschitskii, *Transport policy of Ukraine and its approaching to the norms of the European Union* (Book style). Kiev: Analitic-doradchy center of Blakitnoi strichky, 2010. – 102 p.
- [8] "About priority measures of overcoming the crisis in the transport system of Ukraine" policy brief (Electronic resource).

Visualization of the Ultrasonographical Scans in Critically Ill Neonates

Krcho P., Mihalčo O., Rusnáková S.

Abstract—Ultrasonographical scans and videos in the neonatal period of life are noninvasive possibilities to evaluate some critical information soon after delivery. The aim was to design and test of the innovative application of existing IT solution to new application domains with impact on improved skills in ultrasonography scan interpretation.

Keywords—neonate, ultrasound, visualisation

I. INTRODUCTION

This paper would like to present the process of data storage, evaluation and processing of the ultrasonographical digital pictures and videos in critically ill newborns. During the last years there are increased possibilities to scan different organs and systems bedside (3,4). In the same time, it is important to store data, reevaluate them and use them during the daily rounds and consultations. Early after delivery the ultrasonographical scans are more detailed, because of high extracellular body fluid content. The aims of the ultrasonographical investigation in the newborns are:

1. Discover life threatening conditions, abnormalities, morphologic abnormalities
2. Evaluate the circulation, discover morphologic abnormalities of the heart, evaluate and reevaluate the changes of the circulation based on the transition process of the lung and heart
3. Reevaluate the dynamic changes of the circulation after the established and selected therapeutic interventions
4. Discover and follow closely the changes in different organs
5. Educate nurses and physicians during the daily rounds
6. Project the findings to the big screen for the consultants and specialist during the consultations to decrease the need for reevaluation.

II. ULTRASOUND IN GENERAL

A. Powerfull ultrasound

Clinician performed ultrasound (CPU) is a powerful method for recognizing health and life-threatening conditions of patient, which is given to the hand of caring clinician in every day or night hour. Moreover, CPU often helps clinician to make a therapeutic decision in clinical cases, where there is not enough information gathered from other special examination techniques (invasive arterial blood pressure measurement, invasive pulmonary blood pressure, central venous blood pressure, CT, MRI, ...).

B. Image storage

The more clinicians taking care of the patient, the stronger urge to store images for further revaluation/comparison occurs. Although there are tendentious to unite training in ultrasound, there is still quite a big amount of information bias in description of what is seen on ultrasound screen by different clinicians. This can be prevented by better training program (which is complicated, expensive and time-consuming way), or by storing the images for later

P. Krcho, Neonatal Clinic and Neonatal Intensive Care Unit, Medical faculty of the UPJŠ in Košice, Slovakia (email: peter.krcho@upjs.sk).
O. Mihalčo, Neonatal Clinic and Neonatal Intensive Care Unit, Medical faculty of the UPJŠ in Košice, Slovakia.
S. Rusnáková, Neonatal Clinic and Neonatal Intensive Care Unit, Medical faculty of the UPJŠ in Košice, Slovakia.

inspection/comparison to the up to date finding (which is easier, cheaper and more reliable way).

C. Ultrasound by clinician and technical description of the system.

The system consists of three major components:

1. HiVision Server – is the core part of the system. It is an all-time running service, which accepts and processes store requests from multiple USG devices or other DICOM – compatible devices. The server indexes processed data for faster searching and processing and stores them using local or remote relational database management system. It also publishes web-services in local network, which are used for searching and visualization of previously – stored data.
2. USG device (or other DICOM – compatible devices) are communicating with HiVision Server, communicating and storing data in DICOM format and requesting worklist information. Communication can be achieved by using wired local area network (fixed equipment) or secured wireless network (portable equipment)
3. HiVision Client is user-interface part of the whole system. After authentication and authorization, HiVision Client enables users to look for current and archived patient data, perform postprocessing (i.e. ex-post measurements and image adjustment), categorization, worklist management and other tasks. HiVision Client is available as Windows or Android application. Multiple clients can be connected to single HiVision Server (using published web-services) and share all the data.

Clinician performed ultrasound (or in our case neonatologist performed ultrasound) can be helpful in a great variety of indications and clinical situations. In this paper, we are going to describe three cases, where the storage and later inspection of ultrasound images was helpful in setting the diagnosis, monitoring the development of the finding and setting the prognosis.

III. ULTRASOUND OF THE BRAIN

Neurosonography is a critical part of the care of the sick newborn especially in extremely low birthweight newborns (ELBWN). Sonography is superior to other modalities in imaging of the brain because it can be performed at the bedside, is easily reproducible, and does not require ionizing radiation or sedation. Intraventricular hemorrhage (IVH) in preterm neonates is a devastating consequence of prematurity that has both perinatal and postnatal antecedents [1]. Knowing the timing of IVH is a prerequisite for identifying its antecedents and subsequently applying preventive measures [2,3]. Studies before and after the widespread use of antenatal steroid and surfactant therapy have shown that IVH may occur as early as the 1st min. This observation suggests that IVH may occur in utero, intrapartum or during the early postnatal period [4,5,6]. A large body of evidence has shown that a considerable number of IVH cases occur during the first hours of life [7,8,9,10]. A high percentage of IVH is clinically silent [1,3]. Thus, there is universal consensus that all preterm neonates born at <30 weeks gestation or a birth weight of <1500 g should have a routine screening. Dynamic changes of IVH are shown on Fig.1.

IV. ULTRASOUND OF THE HEART

The role of echocardiography in the neonatal intensive care unit (NICU) has changed over the past few years. Previously, nearly all echocardiography studies in the NICU were performed by pediatric cardiologists to diagnose or monitor congenital heart disease and to screen for patent ductus arteriosus (PDA). More recently neonatologists become interested in the echocardiography assessment of hemodynamic instability in infants.

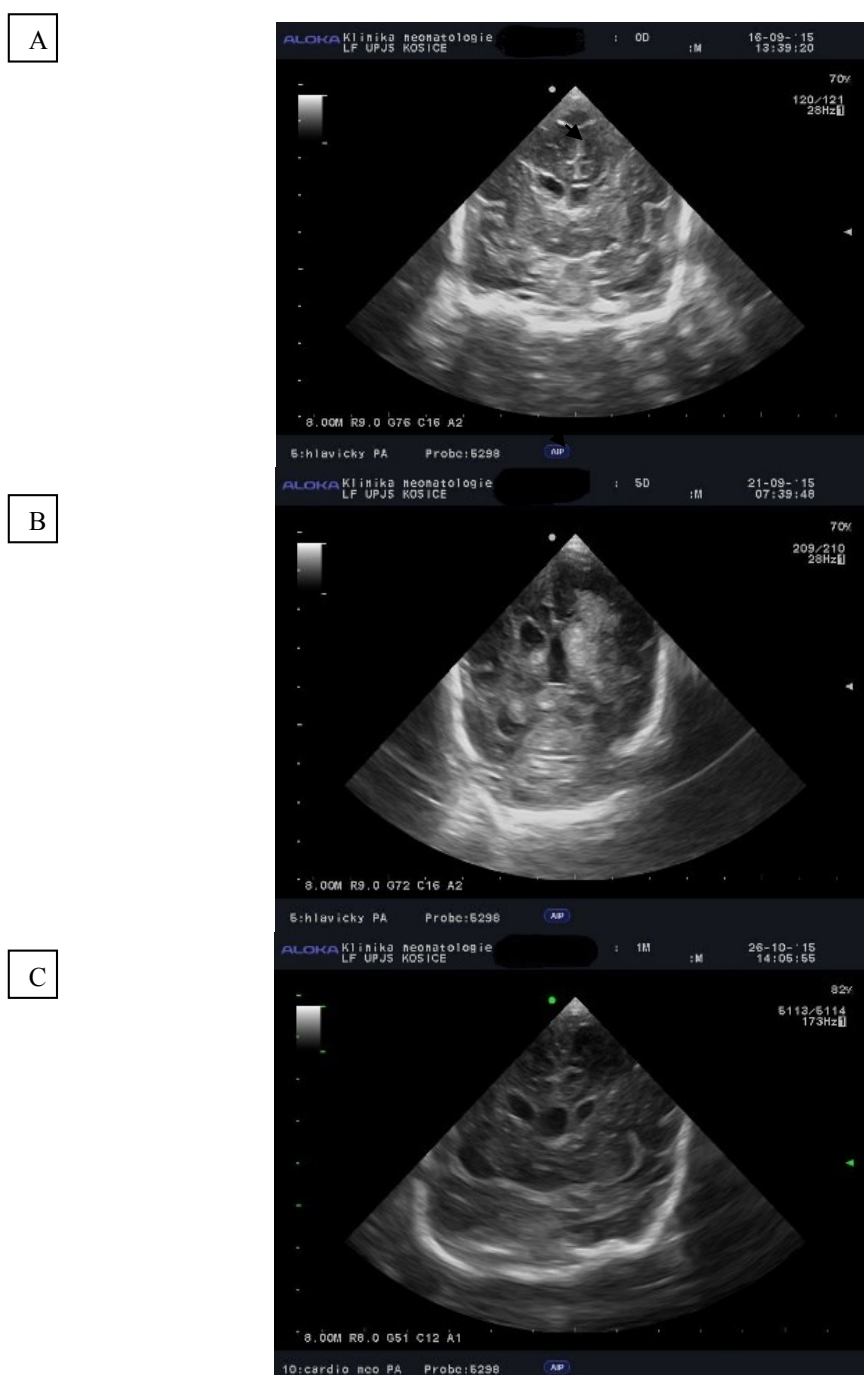


Fig. 1. Morphologic changes of the brain of the ELBWN.

Ultrasound of the brain soon after delivery (A), after the severe intracranial hemorrhage in age of 3 days; black narrow (B) and the control after the healing of the lesion in age of 40 days (C).

A. Assessment of PDA

Although essential for the normal fetal circulation, persistent ductal patency may have significant deleterious effects in preterm or ill term infants. PDAs are found in about half of babies born <29 weeks of gestation and/or weighing <800g [11,12]. Failure of ductal closure, coinciding with the normal postpartum fall in pulmonary vascular resistance, results in left-to-right ductal shunt. The consequences may include pulmonary overcirculation and/or systemic hypoperfusion, both of which may be associated with significant morbidity. Large PDA creates hemodynamic complications including renal insufficiency, necrotizing enterocolitis,

intraventricular hemorrhage, and myocardial ischemia. Correct timing of assessment may lead clinicians to initiate proper treatment, and complications of large PDA could be avoided. Dynamic changes during treatment of PDA are shown on Fig. 2.

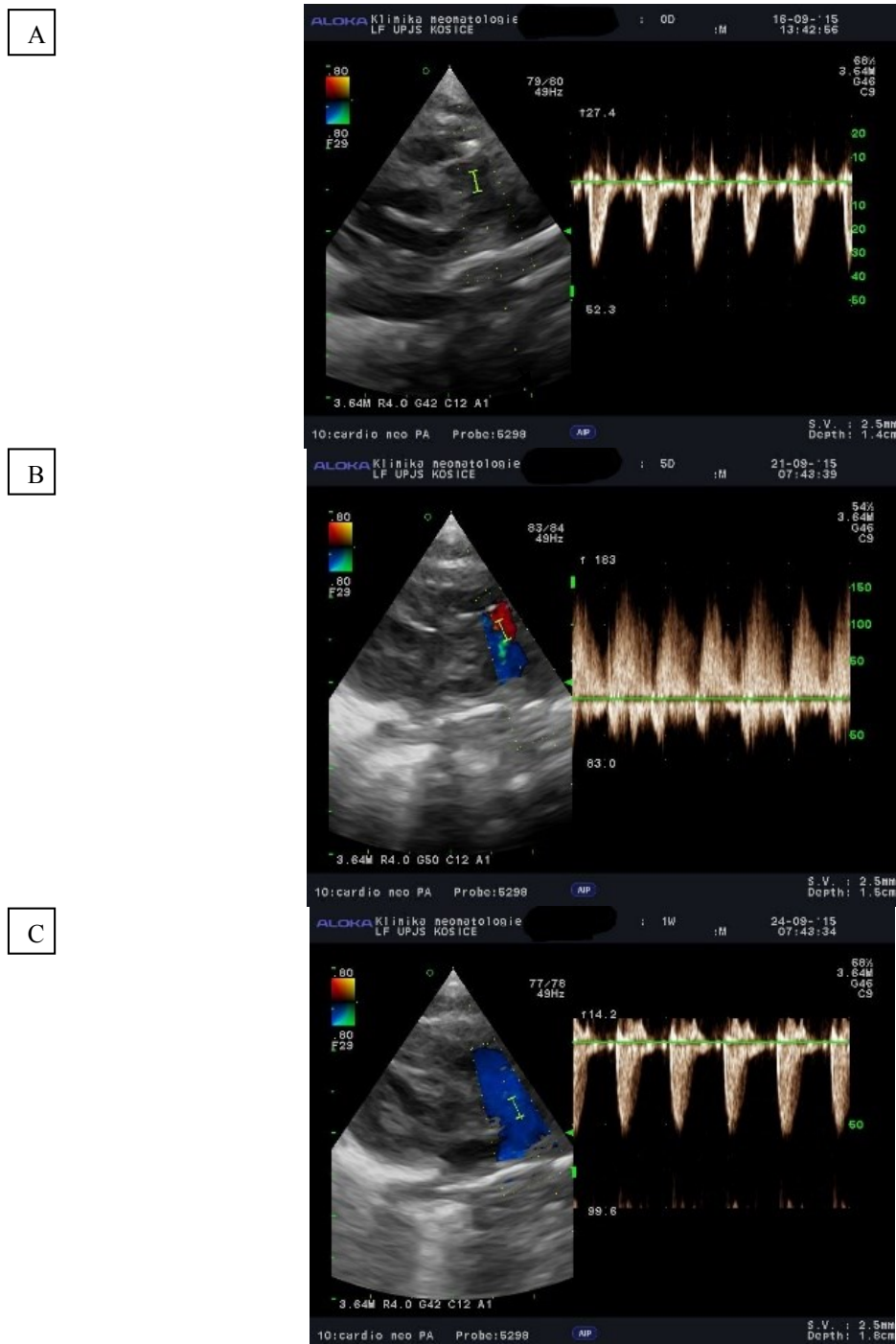


Fig. 2. Doppler flow measurements in the pulmonary artery of the ELBWN. *Soon after delivery (A) , signs of ductus arteriosus persistent in age of 5 days (B) and three days later after the treatment with ibuprofen (C).*

B. Assessment of heart contractility and changes of the cerebral blood flow curves measured by doppler velocimetry.

In severe cases of congenital nonimmune hydrops one time evaluation of the heart contractility, morphology and analysis of the cerebral blood flow measurements could discover

interactions between the heart rate, atrial flutter and changes of the cerebral blood flow curves [14]. This is the reason, why it is important to evaluate the brain and heart together with the changes of the circulation by one person, by neonatologists carrying about the complicated cases. The atrial flutter could be treated by cardioversion with the reevaluation of the cardiac rhythm not only with electrocardiogram, but also with the evaluation of the cerebral blood flow curves giving different view to the pathophysiology of the disease. Fig. 3,4.

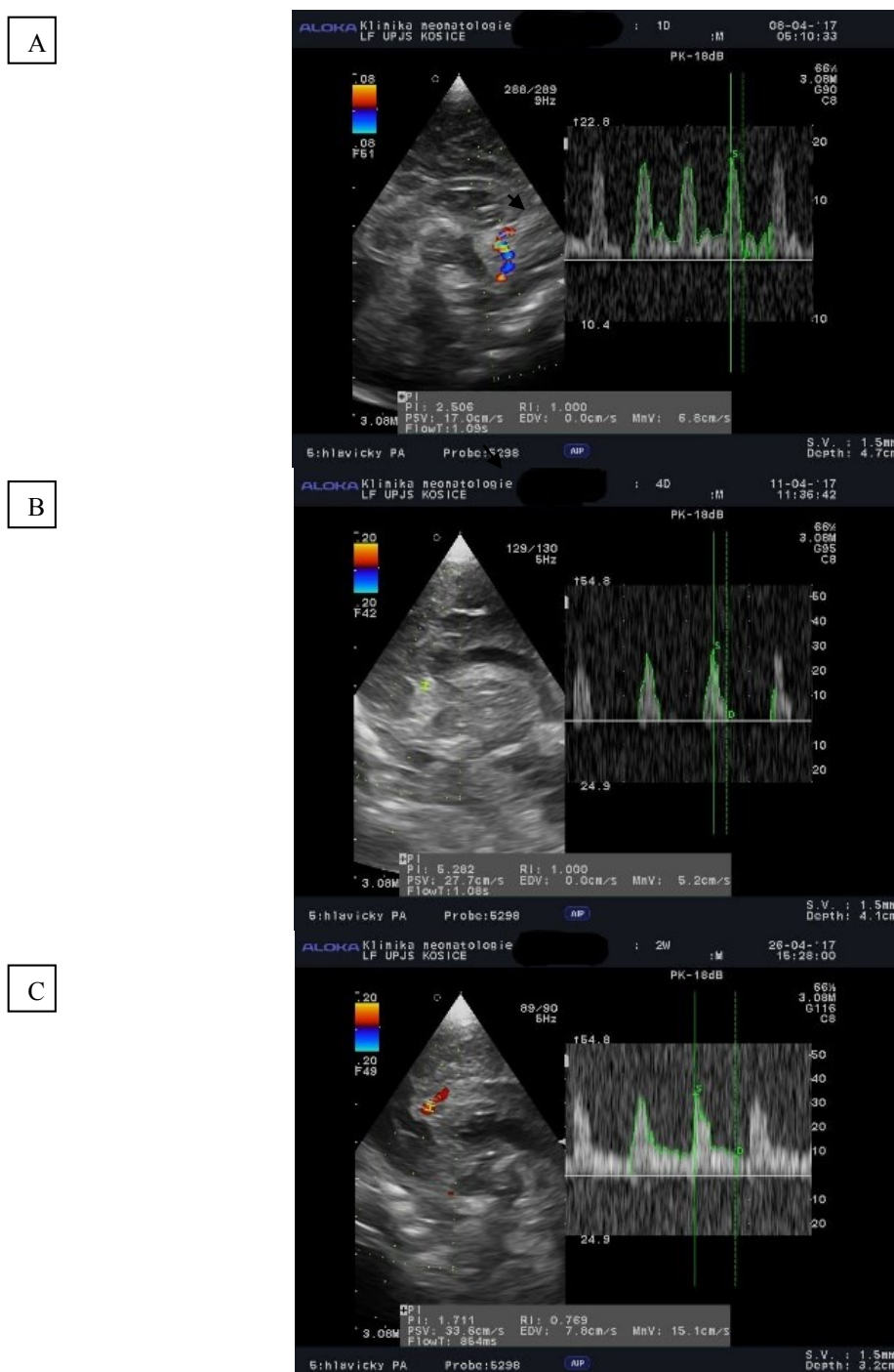


Fig. 3. Changes of the doppler flow curves in the brain.

Changes of the doppler flow velocimetry discovering atrial flutter visible in the diastolic part of the doppler curve (A), pathologic absent diastolic flow (B) and recovery after the treatment (C).

C. Abnormal cardiac rhythms

Fetal arrhythmias are encountered in 1-2% of pregnancies and 10% of these are associated with some form of fetal mortality or morbidity, including structural heart disease, fetal death and neurological complications. The most frequent types of arrhythmia are supraventricular arrhythmias of which the innocent premature atrial depolarizations make up 85%; 10% are tachycardia with a fetal heart rate of over 180/min. [15]

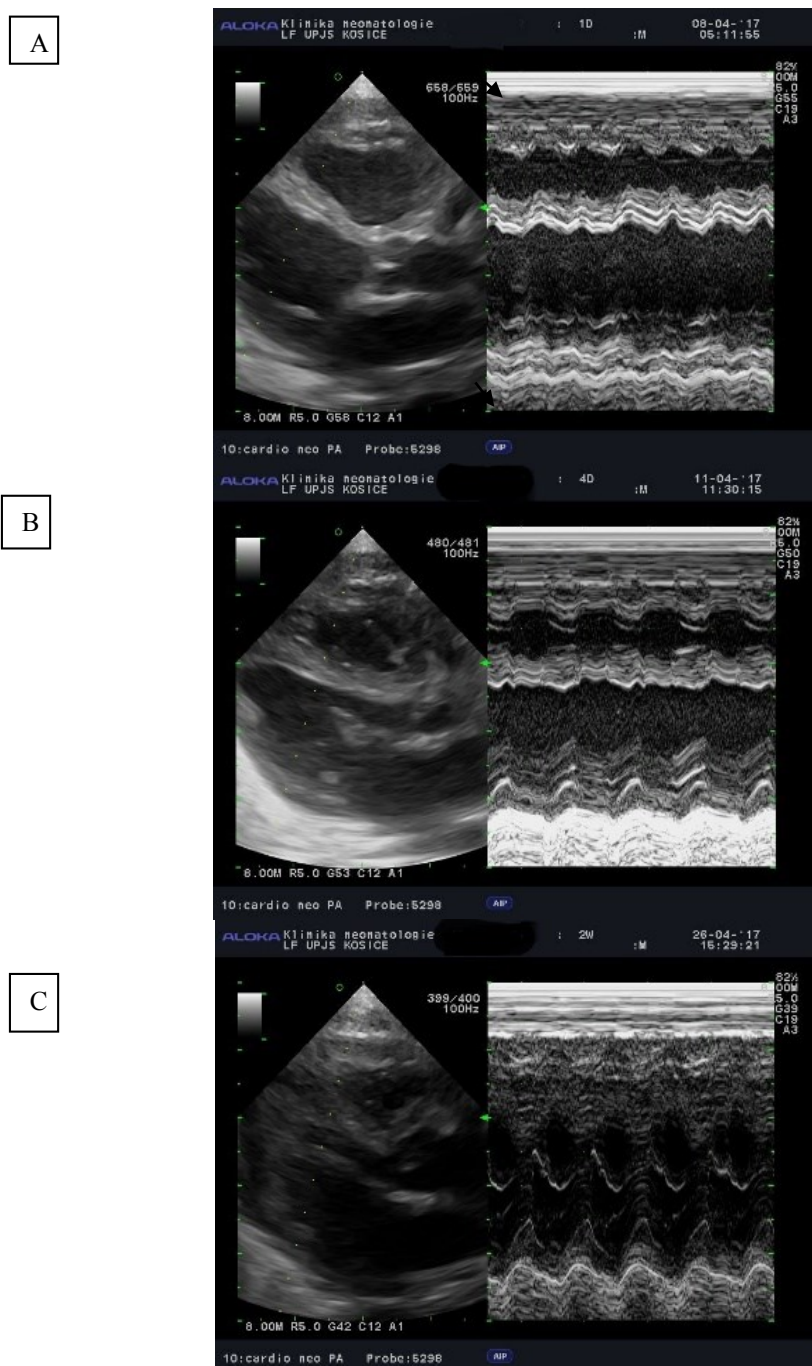


Figure 4. Changes of the cardiac contractility in newborn with congenital hydrops. *Changes of the cardiac contractility (A), after the cardioversion and (B) slow recovery after the inotrope support (C).*

There are data, which suggest that there are autoregulation mechanisms of cerebral blood flow in fetuses with congenital heart disease that enhances cerebral perfusion. [16] Similarly, the autoregulation mechanisms try to maintain stable cerebral blood flow and exchange of metabolically important substances in cerebral circulation of neonate and therefor to protect newborn's brain mainly from hypoperfusion and hypoxia. These mechanisms are effective in the particular range of blood pressure values, below which they become ineffective, causing brain tissue suffering. Brain tissue ultrasound with Doppler measurements of cerebral blood flow is powerful tool, which allows clinician to evaluate the blood flow state in real time and helps to decide if further therapy of cardio-pulmonary system is needed.

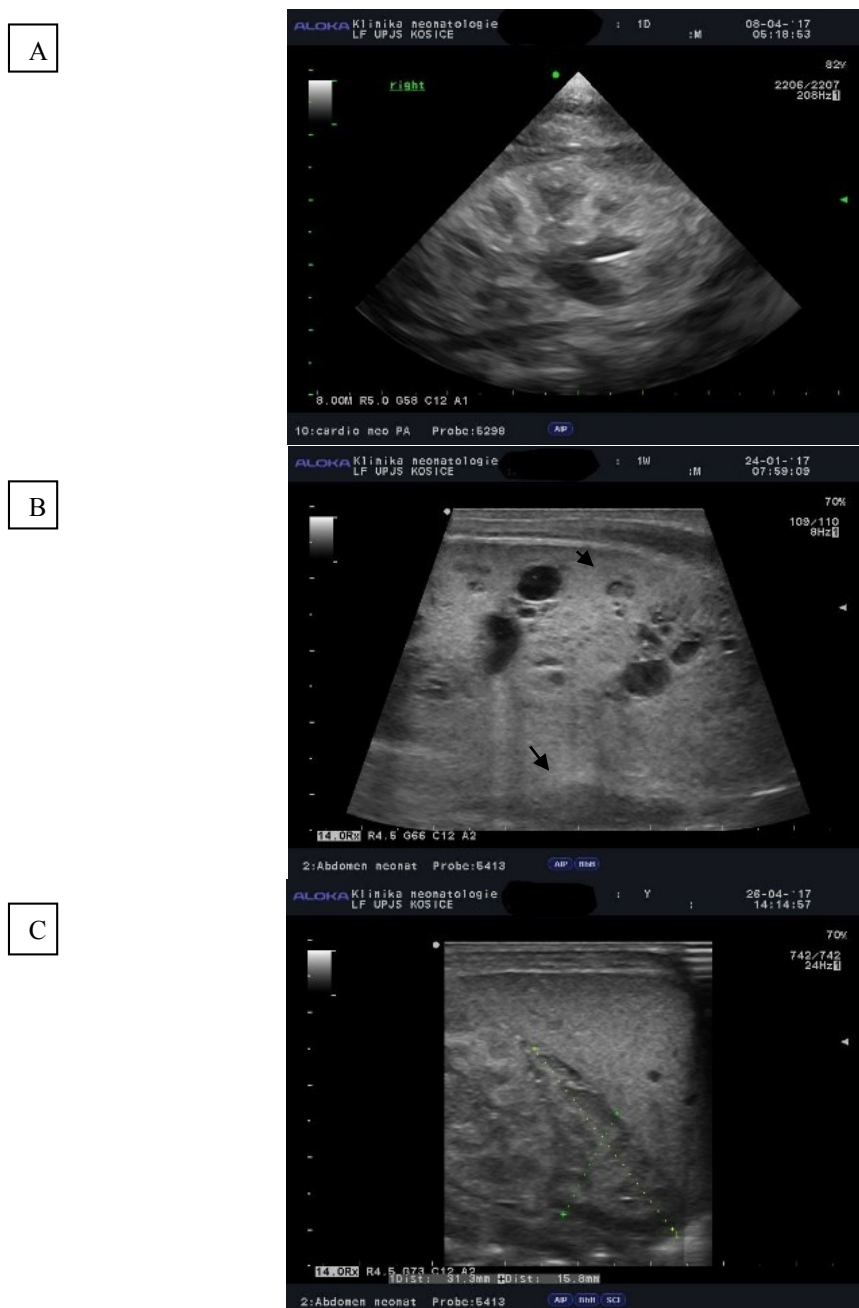


Fig. 5. Ultrasound of the morphologic abnormalities in the kidneys,

Normal kidney (A), autosomal recessive polycystic kidney disease (B) and severe hypertrophy of the suprarenal gland in newborn with congenital adrenal hyperplasia

V. ULTRASOUND OF THE KIDNEY

The ultrasound of the kidney is part of the neonatal screening prenatally and postnatally. The most common lesions are related to the dilatation of the pelvic system. After delivery it is also important to verify the dimensions of the kidneys, discover if there is no agenesis and evaluate the suprarenal glands. Congenital abnormalities of the kidney and urinary tract (ureter, urethra and bladder) are some of the commonest abnormalities identified on ultrasound during pregnancy. Antenatal hydronephrosis is seen in up to 1 in 200 pregnancies. This refers to a dilatation in the collecting system of the kidney that may be physiological, or due to obstruction or reflux. Other possible findings include kidneys that are abnormal in appearance and / or position or bladder abnormalities [16], some of which are shown on Fig. 5.

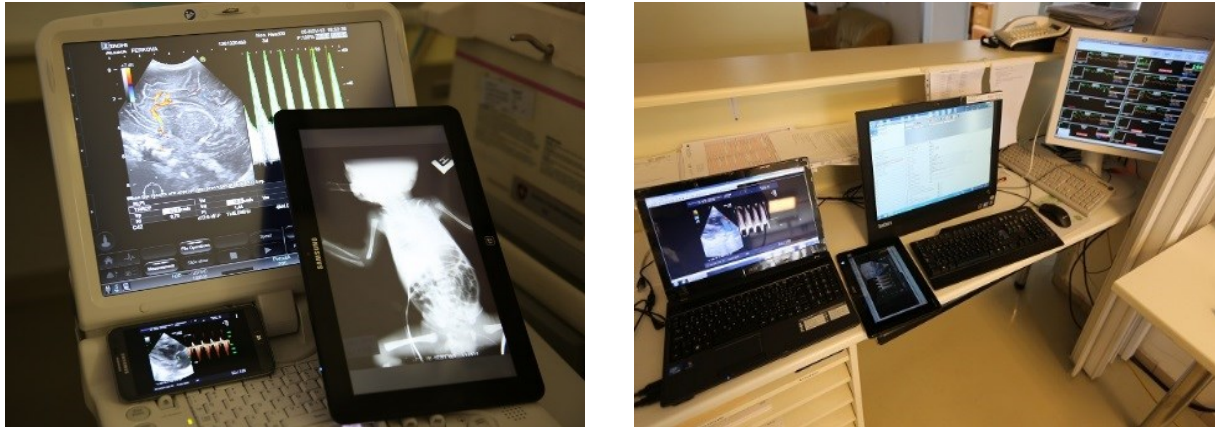


Fig. 6. Hardware used for the data processing

ACKNOWLEDGMENT

The authors presented several clinical situations when bedside ultrasound giving quick information about the changes in the organs giving important information for the clinician bedside. Medical documentation is all over the word based on written data. The physicians trying to classify the disease and select the optimal interventions or treatment. The results of the investigation are printed on papers, recorded or stored in digital media, but the most important is to write the explanation of the scans and findings. It was developed a system enabling the accurate archiving and management of ultrasound images gathered during a clinical screening trial. It is based upon a Windows application utilizing an open-source DICOM image viewer and a relational database. The data process is shared between ultrasound and notebook using secure Wi-Fi. The storage of the scans is done during recording simultaneously in two places.



Figure 7. Projecting data to the big screen during the rounds

The scans are retrieved from the database using tablet and projected to the central digital screen installed closely to the incubators, one big screen for 8 cases. These image records may be processed, stored in a DICOM viewer or export them to external media.

CONCLUSION

The new era need quick evaluation and visualization of the important data. In the neonatal intensive care unit, the newborns are coming in critical condition, the physicians must to act quickly, gently and adequately. The decisions are not easy. The process of the decisions is usually just on written form, the scans are done with x ray, usually one or two pictures, the ultrasound allowing to do several scans from the brain, heart, kidneys, liver, spleen and abdomen. On the other side it is possible to measure several Doppler measurements on the different vessels. It is the huge amount of information, the clinician must to understand them. Our system of the visualization filling several crucial roles in the process:

- Real time data storage
- Discovering crucial abnormalities
- Following closely the decision process of the physician
- Allowing discuss the findings with the consultants and experts
- Enables easy return to the scans done before
- Enables discuss the scans during the daily rounds
- It is used for education purposes for physicians, nurses, students
- Giving a new format for the explanation of the severe abnormalities for parents and caregivers
- Enables monitoring of the treatment process and it's effectivity
- Continual education tool for the staff
- Easy return to the older cases for scientific work and publications.
- Cheap and easy way to work with digital data

REFERENCES

- [1] JJ. Volpe, "Germinal matrix-intraventricular hemorrhage," in *Neurology of the Newborn*. 5th ed. Philadelphia, PA: Saunders Elsevier, 2008 pp. 517–588.
- [2] M. Funato, H. Tamai, K. Noma, T. Kurita, Y. Kajimoto, Y. Yoshioka, "Clinical events in association with timing of intraventricular hemorrhage in preterm infants," *J Pediatr*. 1992, vol. 121, pp. 614–619.
- [3] T. Dolfin, MB. Skidmore, KW. Fong, EM. Hoskins, AT. Shennan, "Incidence, severity, and timing of subependymal and intraventricular hemorrhages in preterm infants born in a perinatal unit as detected by serial real-time ultrasound," in *Pediatrics*, vol. 71, 1983, pp.541–546.
- [4] R. Bejar, V. Curbelo, RW. Coen, G. Leopold, H. James, L. Gluck, "Diagnosis and follow-up of intraventricular and intracerebral hemorrhages by ultrasound studies of infant's brain through the fontanelles and sutures," in *Pediatrics*, vol 66, 1980, pp. 661–673.
- [5] DC. Shaver, HS. Bada, SB. Korones, GD. Anderson, SP. Wong, KL. Arheart, "Early and late intraventricular hemorrhage, the role of obstetric factors," in *Obstet Gynecol*, vol 80, 1992, pp. 831–837.
- [6] S. Sarkar, C. Kaplan, TE. Wiswell, AR. Spitzer, "Histological chorioamnionitis and the risk of early intraventricular hemorrhage in infants born < or = 28 weeks gestation," in *J Perinatol*, vol. 25, 2005, pp. 749–752
- [7] Paneth N, Pinto-Martin J, Gardiner J, Wallenstein S, Katsikiotis V, Hegyi T, et al. Incidence and timing of germinal matrix/intraventricular hemorrhage in low birth weight infants," in *Am J Epidemiol*, vol. 137, 1993, pp.1167–1176.
- [8] P. Emerson, H. Fujimura, P. Howat, D. Howes, J. Keeling, RO. Robinson, "Timing of intraventricular haemorrhage," in *Arch Dis Child*, vol. 52, 1977, pp.183–187.
- [9] GD. Anderson, HS. Bada, DC. Shaver, CJ. Harvey, SB. Korones, SP. Wong, "The effect of cesarean section on intraventricular hemorrhage in the preterm infant," in *Am J Obstet Gynecol*, vol.166, 1992, pp.1091–1099.
- [10] DA. Osborn, N. Evans, M. Kluckow, "Hemodynamic and antecedent risk factors of early and late periventricular/intraventricular hemorrhage in premature infants," in *Pediatrics*, vol. 112, 2003, pp.33–39.

- [11] MD. Reller, MJ. Rice, RW. McDonald, "Review of studies evaluating ductal patency in the premature infants," in *J Pediatr*, vol. 122, 1993, pp. S59–62
- [12] J. Wyllie, "Treatment of patent ductus arteriosus," in *Semin Neonatol*, vol. 8, 2003, pp. 425–432
- [13] M. Szymankiewicz, JE Hodgman, B Siassi, J. Gadzinowski, "Mechanics of breathing after surgical ligation of patent ductus arteriosus in newborns with respiratory distress syndrome," in *Biol Neonate*, vol. 85, 2004, pp. 32–36
- [14] P. Petersen, J. Kastrup, R. Videbrek, G. Boysen, "Cerebral blood flow before and after cardioversion of atrial fibrillation," in *Journal of Cerebral Blood Flow and Metabolism* 9: 1989, pp. 422–425
- [15] Vlagsma, R, Hallensleben E, Meijboom EJ, Supraventricular tachycardia and premature atrial contractions in fetus, *Ned Tijdschr Geneesk*. 2001 Feb 17;145(7):295-9./
- [16] Donofrio, M., Bremer, Y., Schieken, R. et al. *Pediatr Cardiol* (2003) 24: 436.
- [17] Kara, T., Wong, W., Newborn services clinical guideline, online:
<http://www.adhb.govt.nz/newborn/Guidelines/Anomalies/UrinaryTractAnomalies.htm>, 30.04.2017

CERES Press Release

12 July 2017

Co-funded by the
Tempus Programme
of the European Union

Workshop on CERES: Modern Experience on Young Researchers Organization

The workshop CERES: Modern Experience on Young Researcher's Organization was held at University of Žilina on July 5-7. This workshop was organized in frames of conference on Information and Digital Technologies (IDT 2017) in accordance with WP 3.3. *Intense training of teachers and young researchers of TEMPUS CERES project* (ref.no.: 544137-TEMPUS-1-2013-1-SK-TEMPUS-JPHES). Conference on IDT 2017 provides a forum for presentation and discussion of research contributions covering the theories and methods in the field of information and digital technologies, and their application to a wide range of industrial, civil and social sectors and problem areas. IDT 2017 is also an opportunity for teachers, researchers, practitioners, academics and engineers to meet, exchange ideas, and gain insights from each other. IDT 2017 offers a multidisciplinary platform to address the technological, societal and other aspects of information systems.



The main aim of the CERES workshop was to bring together young researches, teachers from academy as well as industry working in all areas of information technologies. We tried to create good atmosphere and conditions for actual and interesting presentations, discussions and cooperation. Our participants have got possibility to joint into other workshops under the conference: (a) New Frontier Information Digital Technology: From Logic Design To Real-World Applications; (b) Dynamical systems and real world applications; (c) Biomedical Technologies and (d) Reliability Technologies.

More than 50 works of workshop participants and experts were presented. Some of presentations of young researchers were discussed at 4 oral workshop's sections and 11 conference's sections. Some of them were included into the well-known IEEE Explorer repository



(<http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=8007434>). It guarantees

CERES Press Release

12 July 2017

Co-funded by the
Tempus Programme
of the European Union

increasing of potential of workshop's participants. In the plenary sessions, distinguished invited experts given a review of the future perspectives in their research areas: reliability, medical informatics, logic design and information technologies. In particular, next invited lectures were been:

- Prof. Frank Coolen, (Durham University, UK) *The Survival Signature for System Reliability*,
- Prof. Radim Bris (Technical University of Ostrava, Czech Republic). Stochastic Ageing and Maintenance Models for Unavailability Quantification of Complex Multi-Component Systems
- Prof. Max A. Viergever (University Medical Center Utrecht, Netherlands). *Challenges in Medical Image Analysis*,
- Prof. Claudio Moraga (Technical University of Dortmund, Germany). *Selected Aspects of Multiple-valued Bent Functions*
- Prof. Charles El-Nouty (Universite Paris 13, France). On the mixed integrated fractional Brownian motion
- Dr. Michal Varga (University of Žilina, Slovakia). Application of diverse techniques and paradigms in pedestrian movement and behaviour simulator, etc.

The organizing team has organized several cultural and social events that were held in pleasant atmosphere. Participants visited to two famous Slovakian castles in Trenčín and Strečno.



The Trenčín castle had seen below its walls hostile armies of Czech and Polish kings, Tatar hordes, imperial landsknechts, soldiers of the Turkish sultan, Tatars of the Khan of Krym, companies of the Duke of Sesia, and Kurucs and Labanec armies. However, it has never been conquered by a direct attack. The oldest stone construction of the castle is the pre-Romanesque rotunda, which dates back to Great Moravia period (at the end of 9th century).

The Strečno Castle is a ruin of a medieval castle of an irregular plan located in northern Slovakia, 12 km east of Žilina. The castle stands on a 103-metre-high cliff above the river Váh. The first recorded mention of the stone castle is from 1316.



CERES Press Release

12 July 2017



Co-funded by the
Tempus Programme
of the European Union



Our team hopes that next workshop on CERES will be organized at the IDT conference too. We sure that this CERES workshop will be good start of participants in their research career. It will give impulse to involve young researchers and to increase their potential in future.

