



# 9<sup>th</sup> Scientific Conference of Young Researchers

## SCYR 2009

**Faculty of Electrical Engineering and Informatics  
Technical University of Košice**

*Proceedings from Conference*

**May 13<sup>th</sup>, 2009  
Košice, Slovakia**



## General partners and sponsors

### General Partner



### Sponsors



**9<sup>th</sup> Scientific Conference of Young Researchers  
of Faculty of Electrical Engineering and Informatics  
Technical University of Košice**

Proceedings from Conference

Published: Faculty of Electrical Engineering and Informatics  
Technical University of Košice  
I. Edition, 279 pages, the number of CD Proceedings: 120 pieces

Editors: prof. Ing. Alena Pietriková, PhD.  
Ing. Attila N.Kovács  
Ing. Jana Modrovičová

**ISBN 978-80-553-0178-5**

**Scientific Committee of 9<sup>th</sup> Scientific Conference of Young Researchers of Faculty of Electrical Engineering and Informatics Technical University of Košice**

Chairman: prof. Ing. Liberios Vokorokos, PhD.

Members:

- prof. Ing. Dušan Kocur, CSc.
- prof. Ing. Dobroslav Kováč, PhD.
- prof. Ing. Irena Kováčová, PhD.
- prof. Ing. Dušan Krokavec, CSc.
- prof. Ing. Dušan Levický, CSc.
- prof. Silaghi Alexandru Marius
- prof. RNDr. Valerie Novitzka, CSc.
- prof. Ing. Alena Pietriková, PhD.
- prof. Ing. RNDr. Ján Turán, DrSc.
- doc. Ing. Roman Cimbala, PhD.
- doc. Ing. Ľubomír Doboš, PhD.
- doc. Ing. Ján Dudáš, DrSc.
- doc. Ing. Žlemíra Ferková, PhD.
- doc. Ing. Jozef Juhár, PhD.
- doc. Ing. Iraida Kolcunová, PhD.
- doc. Ing. Miroslav Mojžiš, CSc.
- doc. Ing. Branislav Sobota, PhD.
- doc. Ing. Iveta Zolotová, PhD.

**Steering Committee of 9<sup>th</sup> Scientific Conference of Young Researchers of Faculty of Electrical Engineering and Informatics Technical University of Košice**

Members:

- Ing. Radoslav Bučko
- Ing. Vierošlava Čáčková
- Ing. Milan Čík
- Ing. Lýdia Dedinská
- Ing. Attila N. Kovács
- Ing. Jana Modrovičová
- Ing. Vladimír Ruščin
- Ing. Igor Vehec, PhD.

**Contact address:**

Faculty of Electrical Engineering and Informatics  
Technical University of Košice  
Letná 9  
042 00 Košice  
Slovak Republic

## Foreword

Dear Colleagues,

research has been one of the main priorities of the university education since its beginning. It is the motto for our Faculty of Electrical Engineering and Informatics too. The ninth Scientific Conference of Young Researchers (SCYR 2009), conference of Graduates and Young researchers, commemorating the 40<sup>th</sup> anniversary of the foundation of Faculty of Electrical Engineering and Informatics, was held at Faculty of Electrical Engineering and Informatics Technical University of Košice on 13<sup>th</sup> May 2009.

Faculty of Electrical Engineering and Informatics has a long tradition of students participating in skilled labor where they have to apply their theoretical knowledge. SCYR is opportunities for doctoral and graduating students use this event to train their scientific knowledge exchange. Nevertheless, the original goal to represent a forum for the exchange of information between young scientists from academic communities on topics related to their experimental and theoretical works in the very wide spread field of electronics, telecommunication, electro-technics, computers and informatics, cybernetics and Artificial intelligence, electric power engineering, remained unchanged.

9<sup>th</sup> Scientific Conference of Young Researchers at Faculty of Electrical Engineering and Informatics Technical University of Košice (SCYR 2009) was traditionally organized in the campus of Technical University of Košice. SCYR 2009 attracted 84 student papers (about 84 participants - young scientists from the Faculty of Electrical Engineering and Informatics) this year. It is an increase comparing to 2008 by 90 % mostly by doctoral categories.

The Conference was opened in the name of dean prof. Ing. Liberios Vokorokos, PhD. by the vicedean of faculty, doc. Ing. Roman Cimbala, PhD. In his introductory address he noted the importance of the Conference as a forum for exchange of information and a medium for broadening the scientific horizons of its participants and stressed the scientific and practical value of investigations being carried out by young researchers.

The program of conferences included two parallel sessions (both consist of oral and poster part):

- Electrical & Electronics Engineering
- Informatics & Telecommunications

with about 84 technical papers dealing with research results obtained mainly in university environment. This day was filled with a lot of interesting scientific discussions among the junior researchers and graduate students, and the representatives of the Faculty of Electrical Engineering and Informatics. This Scientific Network included various research problems and education, communication between young scientists and students, between students and professors. Conference was also a platform for student exchange and a potential starting point for scientific cooperation. The results presented in papers demonstrated that the investigations being conducted by young scientists are making a valuable contribution to the fulfillment of the tasks set for science and technology at Faculty of Electrical Engineering and Informatics.

We want to thank all participants for contributing to these proceedings with their high quality manuscripts. We hope that conference constitutes a platform for a continual dialogue between young scientists.

It is our pleasure and honor to express our gratitude to our sponsors and to all friends, colleagues and committee members who contributed with their ideas, discussions, and sedulous hard work to the success of this event.

We also want to thank our session chairs for their cooperation and dedication throughout the whole conference.

Finally, we want to thank all the attendees of the conference for fruitful discussions and a pleasant stay in our event.

Liberios VOKOROKOS

May 13<sup>th</sup> Košice

## Content

### 1<sup>st</sup> section: Electrical & Electronics Engineering

#### **Michal Aftanas**

*Signal Processing Steps for Objects Imaging Through the Wall with UWB Radar* ..... 14

#### **Vladimír Bánoci, Gabriel Bugár**

*Steganography Systems by using CodeBooks* .....18

#### **Tomáš Béreš, Martin Olejár**

*Bi-directional DC/DC converter controlled by microcontroller* ..... 22

#### **Radovan Blichá**

*Introduction to wireless network and Zigbee*..... 25

#### **Eubomír Cibul'a**

*Direct torque control of asynchronous motor with the help of fuzzy controller*..... 27

#### **Vieroslava Čáčková, Lýdia Dedinská, Milan Kvakovský**

*Micro and Macro Analysis of Polarization Current* ..... 31

#### **Lýdia Dedinská, Vieroslava Čáčková, Milan Kvakovský**

*Dependence of Dissipation Factor on Voltage of Liquid Dielectrics* ..... 34

#### **Juraj Ďurišín**

*Microstructure development of SnAgCu solder joint* ..... 38

#### **Martin Fifik**

*Architecture for Traffic Sign Detection* ..... 41

#### **Daniel Hlubeň**

*Phase-Shifting Transformer in the Electric Power System of the Slovak Republic* ..... 43

#### **Anna Kolesárová**

*Video surveillance systems* ..... 47

#### **Michal Kováč**

*Fuzzy Vector Control of AM* ..... 49

#### **Milan Kvakovský, Vieroslava Čáčková, Lýdia Dedinská**

*Discerption, Effects, Monitoring and Trends of Partial Discharges* ..... 53

#### **František Lizák**

*Basic aspects of smart metering system* ..... 57

#### **Dušan Medved'**

*Numerical Solution of Induction Heating in 2D* ..... 60

<b>Ján Mochnáč, Pavol Kocan</b> <i>Performance evaluation of error resilience and error concealment in H.264</i> .....	64
<b>Ján Molnár</b> <i>Automated oscilloscope measuring</i> .....	66
<b>Maher Nasr</b> <i>Electricity System of Libya and its Future</i> .....	69
<b>Marek Papco, Martin Lojka</b> <i>Adaptation of acoustic models for robust speech recognition</i> .....	72
<b>Peter Poór</b> <i>Bi-parametric model of renewal theory</i> .....	75
<b>Luboš Popovič</b> <i>Modelling of systems with hybrid dynamics</i> .....	79
<b>Jana Rovňáková</b> <i>Complete Signal Processing Procedure for Through Wall Target Tracking: Description and Evaluation on Real Radar Data</i> .....	83
<b>Vladimír Ruščín, Marcel Bodor</b> <i>Soft switching PS-PWM DC/DC converter using auxiliary circuit</i> .....	87
<b>Michal Sakmár</b> <i>Shape errors of generators for ADC testing</i> .....	90
<b>Peter Semančík</b> <i>Thermal degradation of transformer oils</i> .....	92
<b>Ján Šterba</b> <i>Comparison of interpolation methods for estimation of Rayleigh fading channel in OFDM system</i> .....	95
<b>Mária Švecová</b> <i>Taylor Series Based Tracking Algorithm</i> .....	99
<b>Igor Vehec, Pavol Cabúk</b> <i>Methods of Realization of Innerlayer Cavities in Low Temperature Cofired Ceramics</i> .....	103
<b>Tibor Vince</b> <i>Motor Speed Regulation via Internet and Artificial Neural Network</i> .....	107



## 2<sup>nd</sup> section: Informatics & Telecommunications

**Iveta Adamuščinová**

*Knowledge in Software Architectures* ..... 111

**František Baník**

*2D Laser Scanner for the Navigation of the Autonomous Vehicle* ..... 114

**Peter Bratrů**

*Software Maintenance* ..... 117

**Peter Drotár**

*Performance of Orthogonal Space-Time Block Codes* ..... 119

**Zoltán Ďurčák**

*Automatic Web Service Composition* ..... 122

**Juraj Eperješi**

*Map creation based on camera image* ..... 126

**Zlatko Fedor, Tomas Reiff**

*Classification of isolated words with Point-Border Artmap* ..... 130

**Michal Forgáč**

*Adaptive Proposal of Language Modification* ..... 134

**Juraj Gazda**

*Theoretical introduction to nonlinear distorted OFDMA signals* ..... 137

**Daniel Hládek**

*Learning of Fuzzy Rules with Generalization for Dichotomic Classification*..... 141

**Juraj Chovaňák**

*Production lines modeling with the use of Coloured Petri Nets* ..... 145

**Lucia Jancurová**

*Focused magnet for drug targeting* ..... 148

**Lucia Jancurová, Martin Vaľa**

*Data Analysis on Grid using AliEn* ..... 151

**Jozef Janitor, Peter Fecilak**

*Mobility management in VoIP networks:  
Intelligent networks vs. Intelligent applications* ..... 154

**Vladimír Jeleň**

*Image acquisition of cell nuclei in micro-axial tomography* ..... 157

**Marián Jenčík**

*CoAlgebras and Object-Oriented Paradigm* ..... 160

<b>Peter Karch</b> <i>Graph Cut Segmentation</i> .....	162
<b>Ján Kažimír</b> <i>Content management systems using ontology</i> .....	166
<b>Ivan Klimek</b> <i>P2P proxy/cache</i> .....	168
<b>Ján Kliment</b> <i>Mining web logs to improve web site organization and structure</i> .....	173
<b>Pavol Kocan, Ján Mochnáč</b> <i>Mobile Wireless Clients Streaming</i> .....	176
<b>Štefan Köver</b> <i>Fuzzy direct torque control of the AM</i> .....	178
<b>Ján Kunštár</b> <i>Frame Representation of Coherences among UML Model Elements made up of XMI Model Representation</i> .....	182
<b>Miron Kuzma</b> <i>Computational Intelligence in Font Design</i> .....	186
<b>Matej Lakatoš</b> <i>Knowledge about Software Architecture</i> .....	190
<b>Marek Lapko</b> <i>Subgoal Discovery Methods in Reinforcement Learning</i> .....	193
<b>Martin Lojka, Marek Papco</b> <i>Finite-State Machine in Speech Recognition</i> .....	196
<b>Richard Lonščák</b> <i>Design and Application of Optimal Control of Education Model Ball &amp; Plate</i> .....	200
<b>Gabriel Lukáč</b> <i>Opinion mining as yet another approach to information extraction</i> .....	204
<b>Branislav Madoš</b> <i>Design of dataflow computer architecture with tile organization</i> .....	207
<b>Miroslav Michalko</b> <i>Mobile content delivery using Videoserver</i> .....	210
<b>Jana Modrovičová</b> <i>Magnetic Aura of Small Turbojet Engine MPM 20</i> .....	213

<b>Attila N.Kovács, Marek Výrost</b> <i>Basic de/composition of Time Basic Nets</i> .....	217
<b>Ján Papaj</b> <i>Modification of DSR to implement SSV to the Mobile Ad-hoc Network</i> .....	221
<b>Ján Perháč</b> <i>Distributed GPGPU</i> .....	224
<b>Ivan Peťko</b> <i>De/compositional analysis of Petri nets – a survey</i> .....	227
<b>Viliam Ročkai</b> <i>Context for concepts</i> .....	231
<b>Miroslav Sabo</b> <i>Survey on Support for Design Patterns in Software Application Development</i> .....	234
<b>Ján Staš</b> <i>Digital Audio Watermarking in MPEG-1 Audio Layer III: A Survey</i> .....	238
<b>Kristián Šesták</b> <i>Using Virtual Reality Environment for Modeling Software System</i> .....	242
<b>Tamás Tokár</b> <i>Robust video watermarking in DCT domain</i> .....	246
<b>Beáta Tomoriová, Rudolf Andoga, Michal Barto, Norbert Kopčo</b> <i>Contextual adaptation in sound localization: temporal aspects</i> .....	249
<b>Gabriel Tutoky</b> <i>MUDOF Meta-Learning Algorithms for Automatic Selection of Algorithms for Text Classification</i> .....	252
<b>Anita Verbová</b> <i>Linear logic proof search as a stochastic game</i> .....	256
<b>Jozef Vrana</b> <i>Visualization and evaluation of ontological models</i> .....	260
<b>Marek Výrost, Attila N.Kovács</b> <i>On the semantic correspondence of B and BPA specifications</i> .....	263
<b>Marek Vysoký</b> <i>Diagram of security</i> .....	267
<b>Lubomír Wassermann</b> <i>Metalevel Driven Evolution of Software Languages</i> .....	271

**Peter Žársky**

*Information Systems Architectures Driven by Project-knowledge* ..... 275

*Authors's Index* ..... 278

## **1<sup>st</sup> section: Electrical & Electronics Engineering**

# Signal Processing Steps for Objects Imaging Through the Wall with UWB Radar

Michal AFTANAS,

Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

Michal.Aftanas@tuke.sk

**Abstract**—UWB radar system facilitates us to estimate positions and shapes of the objects behind the wall. It has a great utilization especially for rescue and security applications. This paper briefly describes all signal processing steps that are required for imaging of the objects scanned by ultra wideband radar device. It involves the description of the basic scanning method, data preprocessing and calibration, migration methods, wall parameters measurement techniques and compensation of the wall effect on wave that penetrate through the wall. The practical results from real measurements with M-sequence UWB radar device are shown.

**Keywords**—UWB radar system, through the wall imaging, SAR scanning, signal processing.

## I. INTRODUCTION

Ultra WideBand (UWB) electromagnetic waves with frequency's approx. from 0.5 GHz to 3 GHz can penetrate through the non-metallic walls with relatively small attenuation. Such ability is very widely used for the whole field of rescue and security applications. These techniques are most useful when the entering to the room is very dangerous for a man. In such situations, any additional informations about what is currently inside the room and how the room looks like can be helpful for making the strategies before entering the room. Through the wall imaging with UWB radar can be used e.g. to locate hostages or terrorists and weapons behind walls, people trapped in a building during fire, persons buried under fallen walls after earthquake, border controls for the detection of illegal immigrants, cigarettes in trucks, to reconstruct the interior of a room full of smoke during fire, etc.

In this paper, we will refer to the M-sequence UWB radar device used for through the wall scanning, briefly describe whole preprocessing and main processing algorithms. The last part of paper present the practical measurements and results.

## II. M-SEQUENCE UWB RADAR SYSTEM

For measurements presented in this paper the UWB Maximum Length Binary Sequence (M-sequence) radar system [1] [2] is used, because it has many advantages in comparison with classical pulse, or continuous wave radar. The main advantages of UWB radar system are e.g. improved range measurement accuracy and object identification (greater resolution), reduced radar effects due to passive interference (rain, mist, aerosols, metalized strips, ...), decreased detectability by hostile interceptor, availability of low cost transceivers, the UWB signal can be transmitted with no carrier, producing of transmitted signal requires less power, etc. [3].

The first idea to use a very well known M-sequence in UWB radar was proposed in 1996 by Jürgen Sachs and Peter Peyerl,

US patent No. 6272441 [1]. The main advantages of using M-sequence are e.g. the use of periodic signals avoids bias errors, allows linear averaging for noise suppression, M-sequence has low crest factor what allows to use the limited dynamics of real systems and the signal acquisition may be carried out by undersampling. These signals of an extreme bandwidth may be sampled by using low cost, commercial Analog to Digital Converters (ADC) in combination with sampling gates.

The block diagram of M-sequence UWB radar system is shown in Fig. 1. The principle of the M-sequence UWB

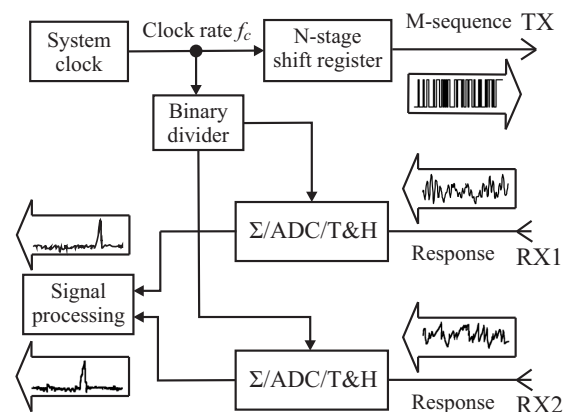


Fig. 1. Block diagram of M-sequence UWB radar system.

radar system can be simply explained as follows. N-stage shift register generates the M-sequence which is transmitted via transmitting antenna, electromagnetic wave is reflected from targets and received via receiving antennas. Received M-sequence is averaged and correlated with transmitted one. The time shift between them corresponds to the distance between transmitter, target and receiver. In principle, the output from M-sequence radar system is the same after the correlation as the output from classical pulse radar system. Therefore, the common preprocessing and imaging algorithms can be used.

## III. SYNTHETIC APERTURE RADAR SCANNING

In order to obtain more information about the investigated object and to narrow antenna flaring angle beam, the Synthetic Aperture Radar (SAR) scanning is applied. The basic 2D SAR spatial model is shown on Fig. 2 a). Transmitted wave is reflected from target to all directions uniformly. Because antenna beam is wide signal reflected from target is received not only when antenna system is exactly over the target, but in all positions that allow to "see" the target. This will cause that pointed target will be represent in acquired B-scan as

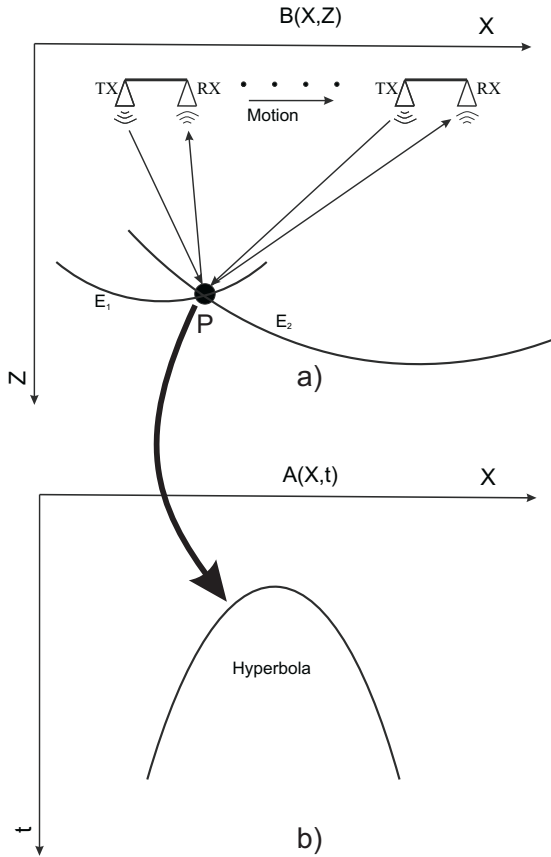


Fig. 2. a) 2D SAR spatial model b) B-scan of pointed target.

hyperbola, like it is shown in Fig. 2 b). Time Of Arrival (TOA) is the time when wave is flying from transmitter to target and back to receiver.

#### IV. CALIBRATION AND PREPROCESSING

Before SAR imaging can be applied to the measured data, several preprocessing and calibration steps have to be undertaken such as: time zero estimation, crosstalk removing, deconvolution and oversampling. Such preprocessing is necessary for imaging of the objects behind the wall and greatly improve the resultant image. Time zero is the time instant in which the transmitted signal leaves the transmit antenna. This time instant has to be shifted at the beginning of the data set for all received impulse responses. Crosstalk is the signal which is transmitted directly from transmitting antenna to the receiving antenna. It does not contain any information about scanned object, but mostly represents the biggest part of received signal. Therefore has to be removed from all impulse responses. The whole system and mostly the antennas have their own impulse responses, which significantly affect the received signal. To reduce this influence, the received impulse responses are deconvolved with the impulse response of the whole radar system including the antennas. Such impulse response of the whole radar system can be measured in anechoic chamber room or in free space towards a big metal plate. The last preprocessing step is to oversample the received impulse responses in the time domain. This step does not improve hardware resolution of the radar system, but can help to find crosstalk and time zero more precisely and can softly improve the image after SAR imaging.

#### V. THROUGH WALL IMAGING ALGORITHM

As it is shown on Fig. 2, the SAR scanning will provide the measured data in  $A(X, t)$  domain. In order to transform  $A(X, t)$  domain back to the  $B(X, Z)$  domain, some migration algorithm have to be used. The signal received in given time can be reflected from all points that lies on the locations where TOA is constant. The points that have the same TOA are on hyperbola  $H$  with focuses at transmitter and receiver positions. This method geometrically focus hyperbolas from  $A(X, t)$  in to the one point in  $B(X, Z)$ .

There are several migration algorithms which can be used to image the objects behind the wall [4]. The simplest imaging method is 2-dimensional SAR migration in time domain [4]. It is a migration with simple geometrical approach often called a back projection [5] or diffraction summation [6] and it does not take into account wave equation. This method is simple to implement, easy to modify, but require big computation power. The similar approach is used in so called Kirchhoff Migration. It is based on solving scalar wave equation. Partial differential equations called separation of variables based on Green's theorem is used to solve this scalar wave equation. Kirchhoff migration theory provides a detailed prescription for computing the amplitude and phase along the wavefront, and in variable velocity, the shape of the wavefront. Kirchhoff theory shows that the summation along the hyperbola must be done with specific weights and, for variable velocity, then the hyperbola is replaced by a more general shape. Kirchhoff migration is mathematically complicated algorithm and is deeply described e.g. in [7], [8]. Wave equation based migration can be done also in frequency domain. Stolt showed that migration problem can be solved by Fourier transform [9]. This process is called f-k migration, or Stolt migration. This method is very fast with low computation complexity, but it is not a very scalable for additional improvements.

Because the antenna flaring angle is not the ideal, the waves that are transmitted or received aslant to the antennas have lower amplitudes. Such signals should be weighted by the antenna footprint function in order to avoid this effect.

#### VI. COMPENSATION OF THE WAVE PENETRATION THROUGH THE WALL

Scanning of the objects behind the wall requires penetrating the electromagnetic wave through the wall. Because the wall has another permittivity, permeability and conductivity as the air, the shape, the amplitude and the velocity of the wave inside the wall will change. The through the wall wave penetrating model is shown in Fig. 3. The conventional method for computing TOA with constant velocity model, which does not consider different velocity in the wall and air, introduces an error in estimation of target shape and position [10]. Therefore, in praxis more or less accurate methods based on ray theory and Snells law are used in algorithms of TOA computation to compensate the presence of the wall [11], [12], [13], [14]. The precise and low complexity algorithm for TOA computation through the wall we described in [15]. This method provides more precise TOA estimation than conventional one and is less complex than three layer methods. Therefore, it is suitable for implementation on realtime hardware.

The wall parameters that are required for TOA estimation such as such wall permittivity and wall thickness are not known a priori and have to be measured. There are several

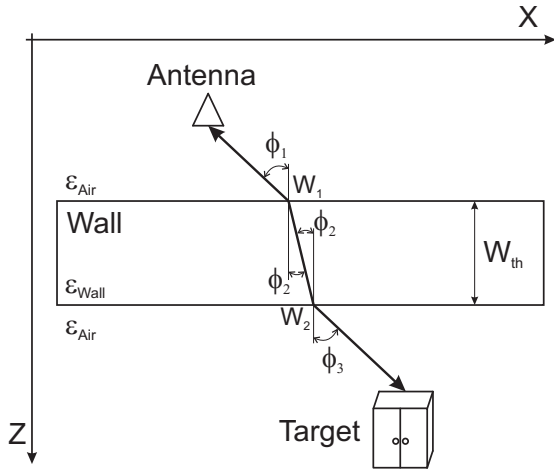


Fig. 3. The model of the wave penetrating through the wall.

methods how to measure them. They can be estimated most precisely when the wall is placed in between the antennas [16], [17]. However this is not practical especially in the case with terrorists or fire since it is meaningless for the intended applications to measure the wall from both sides. A further approach uses different standoff distances for wall parameter estimation in [10], [18]. SAR image de-smearing or auto-focusing can be also used [19], [20]. By representing the wall reflections in the Laplace domain, the pole positions can be used for wall parameter estimation using Prony's method [21]. A model based solution of an inverse problem was also proposed. It solves iteratively the wave equations using Green's function [22], [23]. The time domain reflectometry and the Fresnel equations can be also used [24], [25]

The wave penetrating through the wall is attenuated much more considerably in the wall than it is attenuated in the air. The attenuation inside the wall is mostly depended on the wall conductivity. The magnitude of the wave is reduced with distance even in air. Such attenuation is called spread losses. The spread losses for long distances such as few meters are not neglectable. Because most of the objects that are scanned including walls have flat surfaces, the spread losses can be expressed by reciprocal proportion of distance and wave amplitude. Compensation of the wave attenuation and losses should improve the magnitudes level of all scanned objects according to their reflection properties. However, the small magnitudes from far objects behind the wall are increased including the noise level.

## VII. MEASUREMENTS WITH RADAR DEVICE

For testing the whole string of signal processing steps described above, SAR measurements with 3 scenarios were produced. The measurements were done with M-sequence UWB radar device [1] with frequency range from DC to 2.25 GHz. Two receiver and one transmitter Horn antennas with frequency range approx. 0.5 GHz to 4 GHz were used. The measurement description with scanned results of scenario 1 is shown in Fig. 4 [26]. The aquarium filled with water was placed behind the 18 cm thick brick wall. It can be seen, that the position of the aquarium is shown at correct position, because the algorithm for precise TOA computation described above was used.

The measurement description with scanned results of scenario 1 is shown in Fig. 5 and Fig. 6 respectively [15]. The

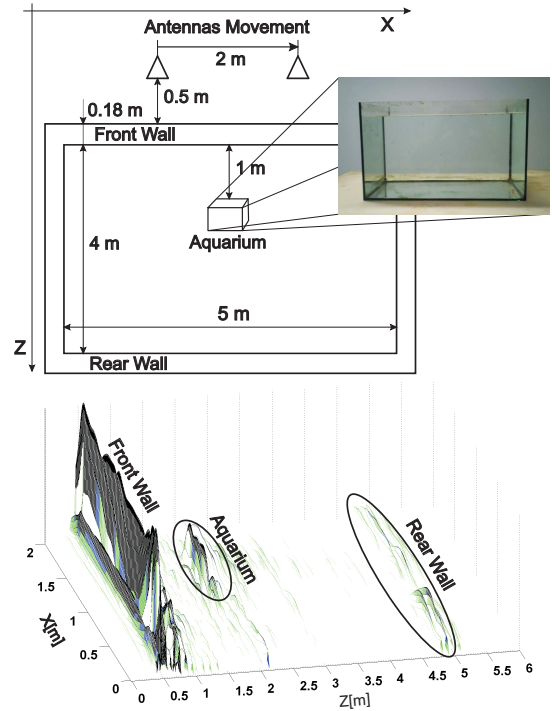


Fig. 4. SAR measurement, scenario 1 resulting in correct target positions after respecting wave propagation in the wall.

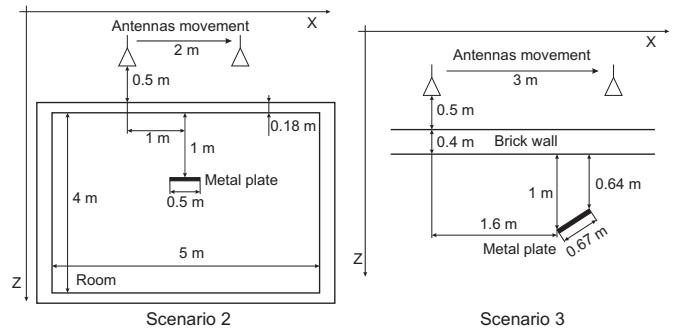


Fig. 5. SAR measurement, scenario 2 (left), scenario 3 (right).

thin metal plate was placed behind the 18 cm and 40 cm thick brick walls. It can be seen, that the correct position of metal plate can be obtained only when the precise method of TOA computation is used Fig. 6 a), b). For the metal plate that is placed aslant to the wall a simple compensation algorithm described in [15] is not sufficient for obtaining the correct results Fig. 6 c), d). The precise method of TOA computation described above have to be used.

## VIII. CONCLUSION

In this paper the signal processing steps that are required for objects imaging through the wall with UWB radar were described. Theoretical approach of SAR scanning, data preprocessing, calibration, migration and compensation of wall effect were tested on real measurements with M-sequence UWB radar device. Three scenarios were measured and processed. From shown results can be seen, that the proposed algorithms and radar device are suitable for imaging of the objects behind the wall. The whole system can be used for rescue and security applications.



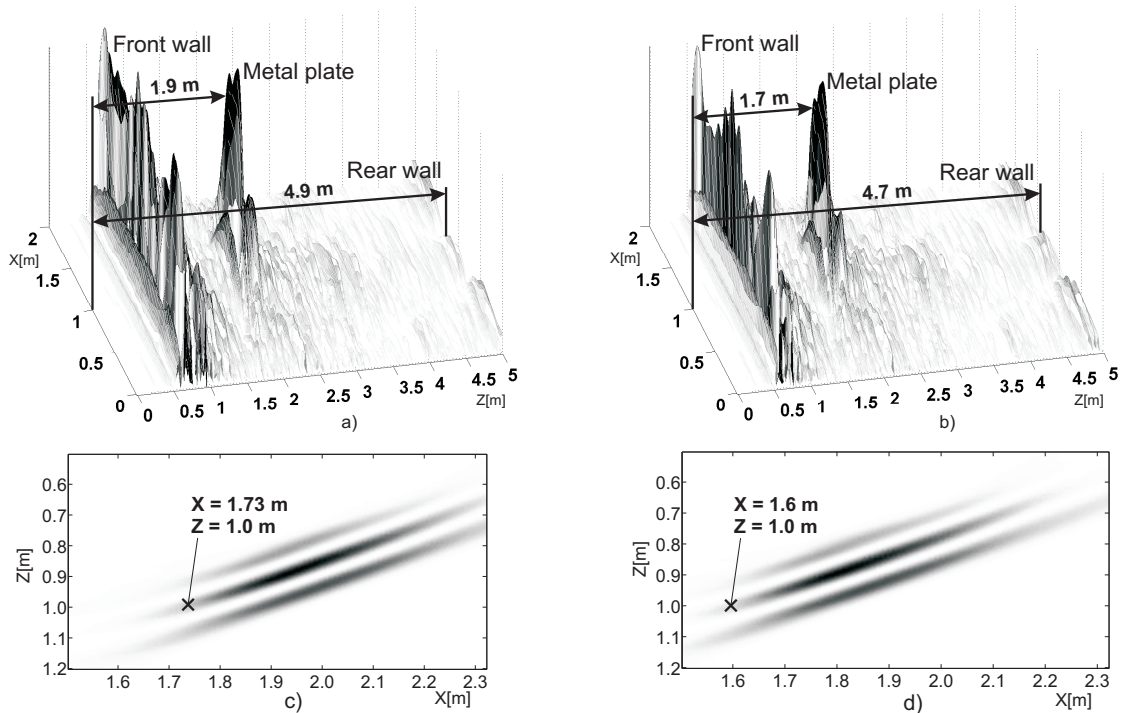


Fig. 6. Scenario 2: a) without wall compensation, b) with proposed wall compensation. Scenario 3: c) with simple wall compensation, d) with proposed wall compensation.

#### ACKNOWLEDGMENT

This work was supported by the European Commission through the 6th framework under the contract COOP-CT-2006-032744 RADIOTECH and the Slovak Research and Development Agency under the contract No. LPP-0287-06.

#### REFERENCES

- [1] D. Daniels, "M-sequence radar," in *Ground Penetrating Radar*. London, United Kingdom: The Institution of Electrical Engineers, 2004.
- [2] J. Sachs, P. Peyrerl, S. Wockel, M. Knec, R. Herrmann, and R. Zetik, "Liquid and moisture sensing by ultra-wideband pseudo-noise sequence signals," *Meas. Sci. Technol.*, pp. 1074–1087, Apr. 2007.
- [3] I. Immoreev and Fedotov, "Ultra wideband radar systems: advantages and disadvantages," *Ultra Wideband Systems and Technologies*, pp. 201–205, 2002.
- [4] M. Aftanas, "Through Wall Imaging Using M-sequence UWB Radar System," Thesis to the dissertation examination, Technical University of Kosice, Department of Electronics and Multimedia Communications, Slovak Republic, Feb. 2008.
- [5] L. Ulander, H. Hellsten, and G. Stenstrom, "Synthetic aperture radar processing using fast factorized back-projection," *Aerospace and Electronic Systems*, vol. 39, pp. 760–776, July 2003.
- [6] D. Miller, M. Oristaglio, and G. Beylkin, "A new slant on seismic imaging," *Migration and integral geometry: Geophysics*, vol. 52, pp. 943–964, July 1987.
- [7] G. F. Margrave, "Numerical methods of exploration seismology with algorithms in matlab," Master's thesis, Department of Geology and Geophysics, The University of Calgary, Jan. 2001.
- [8] O. Yilmaz, *Seismic Data Processing, volume 2 of Investigations in Geophysics*. Tulsa: Society of Exploration Geophysicists, Oct. 1987.
- [9] R. H. Stolt, "Migration by Fourier transform," *Geophysics*, vol. 43, pp. 23–48, Feb. 1978.
- [10] G. Wang and M. G. Amin, "Imaging Through Unknown Walls Using Different Standoff Distances," *Signal Processing, IEEE*, vol. 54, pp. 4015–4025, Oct. 2006.
- [11] S. Gauthier, E. Hung, and W. Chamma, "Surveillance Through Concrete Walls," *Proceedings of SPIE 5403*, pp. 597–608, Dec. 2004.
- [12] F. Ahmad and M. G. Amin, "High-Resolution Imaging using Capon Beamformers for Urban Sensing Applications," *Acoustics, Speech and Signal Processing, ICASSP, IEEE*, vol. 2, pp. 985–988, Apr. 2007.
- [13] C. Lei and S. Ouyang, "Through-wall Surveillance using Ultra-wideband Short Pulse Radar: Numerical Simulation," *Industrial Electronics and Applications, 2007*, pp. 1551–1554, May 2007.
- [14] F. Ahmad, M. G. Amin, and S. A. Kassam, "Synthetic aperture beam-former for imaging through a dielectric wall," *Aerospace and Electronic Systems, IEEE*, vol. 41, pp. 271–283, Jan. 2005.
- [15] M. Aftanas, J. Rovnakova, M. Drutarovsky, and D. Kocur, "Efficient Method of TOA Estimation for Through Wall Imaging by UWB Radar," *IEEE International Conference on Ultra-Wideband (ICUWB2008)*, vol. 10, pp. 101–104, Sept. 2008.
- [16] A. Muqabel and A. Safaai-Jazi, "A new formulation for characterization of materials based on measured insertion transfer function," *Microwave Theory and Techniques*, vol. 51, pp. 1946–1951, Aug. 2003.
- [17] G. Cui, L. Kong, J. Yang, and X. Wang, "A New Wall Compensation Algorithm for Through-the-wall Radar Imaging," *Synthetic Aperture Radar, 2007. APSAR 2007*, pp. 393–396, Nov. 2007.
- [18] H. Wang, Z. Zhou, and L. Kong, "Wall Parameters Estimation for Moving Target Localization with Through-the-Wall Radar," *Microwave and Millimeter Wave Technology, 2007*, pp. 1–4, Apr. 2007.
- [19] F. Ahmad, M. G. Amin, and G. Mandapati, "Autofocusing of through-the-wall radar imagery under unknown wall characteristics," *IEEE transactions on image processing*, vol. 16, pp. 1785–1795, July 2007.
- [20] G. Mandapati and M. Amin, "Blurriness and focusing-defocusing for Through the Wall Radar Imaging," *Signal Processing and Information Technology, 2004*, pp. 246–249, Dec. 2004.
- [21] A. T. S. Ho, W. H. Tham, and K. S. Low, "Improving classification accuracy in through-wall radar imaging using hybrid prony's and singular value decomposition method," *Geoscience and Remote Sensing Symposium, 2005*, vol. 6, pp. 4267–4270, July 2005.
- [22] R. Linnehan, J. Schindler, D. Brady, R. Kozma, R. Deming, and L. Perlovsky, "Dynamic Logic Applied to SAR Data for Parameter Estimation Behind Walls," *Radar Conference, 2007 IEEE*, pp. 850–855, Apr. 2007.
- [23] C. L. Bastard, V. Baltazart, Y. Wang, and J. Saillard, "Thin-Pavement Thickness Estimation Using GPR With High-Resolution and Superresolution Methods," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, pp. 2511–2519, Aug. 2007.
- [24] F. Sagnard and G. E. Zein, "In Situ Characterization of Building Materials for Propagation Modeling: Frequency and Time Responses," *IEEE Transactions on Antennas and Propagation*, vol. 53, pp. 3166–3173, Oct. 2005.
- [25] H. Khatri and C. Le, "Identification of Electromagnetic Parameters of a Wall and Determination of Radar Signal Level Behind a Wall," *Proceedings of SPIE, the International Society for Optical Engineering*, vol. 6210, pp. 62 100Q.1–62 100Q.7, Apr. 2006.
- [26] J. Sachs, M. Aftanas, S. Crabbe, M. Drutarovsky, R. Klukas, D. Kocur, T. Nguyen, P. Peyrerl, J. Rovnakova, and E. Zaikov, "Detection and Tracking of Moving or Trapped People Hidden by Obstacles using Ultra-Wideband Pseudo-Noise Radar," *5th European Radar Conference, EuRAD 2008*, pp. 408–411, Oct. 2008.

# Steganography Systems by using CodeBook

<sup>1</sup>Vladimír BÁNOCI, <sup>2</sup>Gabriel BUGÁR

Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

<sup>1</sup>vladimir.banoci@tuke.sk, <sup>2</sup>gabriel.bugar@tuke.sk

**Abstract**—In this paper, we present a novel method called Code Book in steganography system based on CDMA (Code Division Multiple Access) techniques having regard to perceptibility of stego-image. These techniques and their variations are widely used in radio telecommunication systems. As it will be shown later, apply imposed features of CDMA techniques are in consonance with imperatives claimed upon steganography systems.

**Keywords**—Steganography, Discrete Cosine Transformation (DCT), Code Book, CDMA, PN sequence

## I. INTRODUCTION

Steganography is a hiding technique that has been mainly used in information applications. The basic principle lies in embedding the secret message into a camouflage media to ensure that an unintended party will not be aware of the existence of the embedded secret in stego-media hence its goal is to hide the presence of communication [1], [2].

Images are the most popular cover media for steganography. A popular digital steganography technique is so-called Least Significant Bit (LSB) embedding. With the LSB embedding technique, the two parties in communication share a private secret key that creates a random sequence of samples of a digital signal. The secret message, possibly encrypted, is embedded in the LSBs of those samples of the sequence [3].

Steganalysis is the science of detecting hidden information. The main objective of steganalysis is to break steganography system. There are three main types of steganalysis [1], [4]:

(a) Visual attacks try to reveal the presence of hidden information through inspection with the naked eye or with the assistance of a computer, which can separate the image into bit planes for further analysis.

(b) Statistical attacks are more powerful and successful, because they reveal the smallest alterations in an images statistical behavior. These attacks can be further classified as (i) Passive attack and (ii) Active attack. Passive attacks deal with identifying the presence or absence of a covert message or the embedding algorithm used etc. whereas the goal of active attacks is to estimate the embedded message length or the locations of the hidden message or the secret key used in embedding [4].

(c) Structural attacks are based on fact that the format of the data files often changes as the data to be hidden is embedded; on identifying these characteristic structure changes can detect the existence of image.

For such reasons, many imagery steganographic methods have been invented. Here we briefly review research carried out particularly in the Discrete Cosine Transformation (DCT) domain as JSteg method that hides information sequentially in LSBs of the quantized DCT coefficients (qDCTCs) while

skipping 0s and 1s [5]; OutGuess method scatters information into the LSB of qDCTCs [6]. Embedding is followed by a correction procedure to ensure that the distributions of any related pair of the qDCTCs are unchanged. Another method employs the technique of matrix encoding to hold secret information using LSB of qDCTCs in F5 [7]. Others researches of using DCT transformation are mentioned in [8], [9].

Steganography channel described in this paper is represented by hiding secret information - image in cover image. Conventional steganographic models use the knowledge of cover-object and stego-object to extract embedded data. The embedding process of designed method is carried out by hiding secret image in each block of quantized DCT coefficients.

The acquisition of proposed methods using CDMA techniques is enhancement of security level in steganography systems. Firstly, better distortion of secret message, in this manner, the stego-image quality degradation is more imperceptible to the human eye. Secondly, adding the cryptographic element to steganographic algorithm in process of embedding, is represented by characteristics of the CDMA. Moreover, CDMA mathematical apparatus applied in our method is instrumental in extraction of secret message ergo it allows lower embedding energy of secret message. Experimental results have demonstrated that applied CDMA has its contribution in hiding communication of steganography system.

## II. SIMULATION OF HUMAN PERCEPTION

The basic principle of embedding secret object is defined in sense of similarities stego and cover object and its measure can be expressed by function of conformity. However, this function of perceptibility does not have practical use. Hence, the objective measures, which stem from statistical approach, are generally used in practice. Their application is based on measuring the distortion between cover data  $f(i, j)$  with resolving capacity  $M \times N$  of picture element and modified data  $f'(i, j)$  with same resolving capacity [10]. To the most frequent used criterions of evaluating the quality of reconstructed information belong Mean Square Error (MSE), its normalize value NMSE; Signal Noise Ratio (SNR) and Peak Signal Noise Ratio (PSNR). The PSNR value is used in all our practical examples as mathematical simulation model of human visual perception, given by following way:

$$PSNR = 10 \log_{10} \frac{(2^n - 1)^2}{MSE} [dB] \quad (1)$$

$$MSE = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N [f'(i, j) - f(i, j)]^2 \quad (2)$$

where the variable  $n$  represents the brightness resolution (bit/pixel).

### III. DISCRETE COSINES TRANSFORMATION

It transforms signal or image from spatial domain to frequency domain. In this method all experiments have been doing with 2D-DCT. The 2-D basis functions can be generated by multiplying the horizontally oriented 1-D basis functions with vertically oriented set of the same functions, therefore the 2D-DCT is a direct extension of the 1-D case and is given by the following way:

$$F(u, v) = \left(\frac{2}{N}\right)^{\frac{1}{2}} \left(\frac{2}{M}\right)^{\frac{1}{2}} \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} C(i)C(j) \cos\left(\frac{(2i+1)u\pi}{2N}\right) \cos\left(\frac{(2j+1)v\pi}{2M}\right) f(i, j) \quad (3)$$

Variable  $F(u, v)$  is block of transformed coefficients of input picture elements  $f(i, j)$ . Variables  $C(i)$  and  $C(j)$  are given by:

$$C(\gamma) = \begin{cases} \frac{1}{\sqrt{2}} & \text{for } \gamma = 0 \\ 1 & \text{otherwise} \end{cases} \quad (4)$$

Energy of transformed frequency domain is centered into coefficients with lower spatial frequencies. The most energy is embodied in coefficient with zero spatial frequency DC coefficient. The energy is decreasing towards lower frequencies, what usually depends on nature of the cover image. Each DCT transformation coefficient can bear information equal 1 bit of secret message. If more coefficients are affected in process of embedding secret message, more distortion is generated, what also results to higher detectability by attacker. This approach underlies the DCT method that is used as reference method for comparison to our proposed methods.

The advantage of DCT transformation compared to more efficient transformation Karhunen-Loeve Transform (KLT), which is optimal in the sense of energy compaction, is DCT transformation that can be partially precomputed so faster implementation is possible in embedding algorithm [9]. In addition to Discrete Fourier Transformation (DFT), studies have shown that DCT provides better energy compaction than DFT for most natural images [10].

### IV. CODE DIVISION MULTIPLE ACCESS (CDMA)

The CDMA technique emanates from Direct - Sequence Spread Spectrum (DSSS) approach that stands on mathematical apparatus of orthogonal pseudorandom number sequences (PN Sequence) which allow multiple access on the same medium without interference. This fact is achieved by orthogonality and autocorrelation phenomenon of PN sequences, where autocorrelation represses the influencing of noise on steganographic channel [11].

PN sequence is converted to the states -1 and 1, consequently multiplied by information bit, as it is shown in Fig.1 Hence, PN sequence is summed up with other PN sequences carrying their own information bit. However, this technique of spreading information cannot be considered as vital to the capacity of steganography channel. In spite of, this spreading technique (DSSS) by PN sequences incites to use the other attributes applicable in steganography. Firstly, PN sequences,

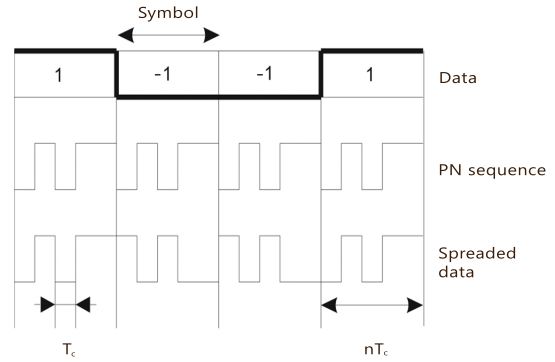


Fig. 1. The explanation of spreading information

depending on its type, have subequal quantity of -1 and 1, thus secret data do not directly figure in bulks or other periodical sequences, which can be an incitement to attacker about concealed communication. Secondly, spreading information bit by PN sequence allows using its correlation characteristic. The longer length of PN sequence, better correlation results can be achieved, therefore higher rate of defective bits during the extraction can be admitted

Needless to say, applying longer PN sequence claims to higher capacity of steganography channel with same secret image. This approach finds the balance between several steganography factors as capacity, modest computing power, undetectability and altogether unsusceptibility of communication.

### V. DESIGN OF PROPOSED ALGORITHM

The proposed method *CodeBook* stands on dividing input stream of secret message to binary code words, where the choice of its length  $d$  ( $d \geq 2$ ) affects spreading technique. There is the unique PN sequence dedicated to each state of code word. Deliberating this fact, the length of code word defines the number of PN sequences used in spreading process, which is given by following  $2^d$ ; where  $d$  is length of code word. The length of PN sequence linearly rises with code length  $d$  in dependence on type of PN sequence generator [12]. For our purpose were selected msequences, with good autocorrelation attributes and the choice of its length is given by level of generating multinomial. Thereafter, this method Code Book is represented by two columns, where on the left side are all states of code word of the length -  $d$  and on the right side are associated PN sequences to these rows ergo words. The open variability of associating PN sequences to code words represents additional restrains to attacker to detect communication in steganographic channel. In addition, the using of PN sequence in this method also arises eventuality of application CDMA technique. Therefore, our proposed method was enhanced by multi - access approach derived from CDMA systems, where our goal was to increase capacity with comparably same perception results. The modified method Code Book with CDMA is represented by code division double access, where one PN sequence from code book is multiplied by 1 and another PN sequence is multiplied by -1. These information bits 1 and -1 denote position in which PN sequence were embedded in the steganographic channel. Hence it follows, summation of these PN sequences, where each represents one code word, means double compression

of information stream needed to transfer in contrast to Code Book method.

## VI. RESULTS

The comparison of capacity between the reference DCT method, Code Book and CDMA Code Book method is shown in the Tab.II and Fig.2. The capacity of particular method was calculated depending on offset variable, which denotes number of DCT coefficient induced in descending order of its frequencies. During the practical experiment the offset value is preset to offset=8 what practically means that first seven DCT coefficient are not used in process of embedding.

TABLE I  
THE CAPACITY (*in Bytes*) OF THE STEGO-IMAGES CREATED BY THE PROPOSED METHOD USING VARIOUS LENGTH OF CODE WORD

d	Embedding algorithm		
	DCT Method	Code Book	Code Book with CDMA
2	7296	4864	9728
3	7296	3126	6254
4	7296	1945	3891

The relation between the length of code word used in proposed methods and PSNR ratio with 100% of successfully extracted secret message depicts in both, Tab.II 2 and Fig.3, where letter d represents the length of code word from Code Book. The embedding picture in mentioned experiment results is win40x40.bmp (1602 B) and cover image is lena256x256.bmp (66 614 B).

TABLE II  
THE PSNR VALUE (*in dB*) OF THE STEGO-IMAGES CREATED BY PROPOSED METHOD USING VARIOUS LENGTH CODE.

d	Embedding algorithm		
	DCT Method	Code Book	Code Book with CDMA
2	48,62	54,02	49,73
3	48,62	54,29	50,02
4	48,62	54,78	53,80

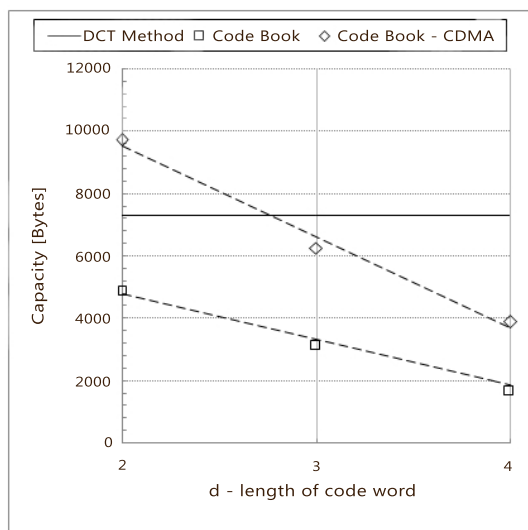


Fig. 2. The capacity of referenced and proposed methods

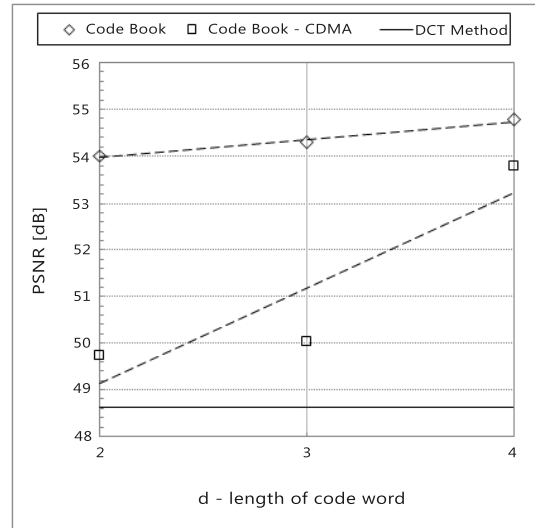


Fig. 3. The comparison of PSNR value (in dB) of the stego-images created by proposed method using various length of code word

## VII. DISCUSSION

Even though the capacity of Code Book method is smaller than DCT reference method, the practical results shown in exhibits point out that applying PN sequences from Code Book have better PSNR ratio than DCT method. Hence, the proposed Code Book with CDMA method was designed for purpose of increasing capacity of the Code Book method. Nevertheless, this aim has been accomplished partially. As the Fig.2 demonstrates the Code Book with CDMA has better capacity results than Code Book, even higher than DCT method in case of using code word length of 2, what practically means compression of the source information. However, autocorrelation function does not perform such a good results comparing to Code Book method with shorter length of code word. Also, from the figure it is obvious that scaling up of word length is followed with better correlation results, therefore better decomposition of source information can be achieved. In consideration of this fact, Code Book with CDMA method is more sensitive to noise in communication channel rather than Code Book method.

## VIII. CONCLUSION

There are two criteria considered according to steganography imperatives. The first mentioned method Code Book improves the perceptibility of stego-images, where capacity does not play the main role. The second method Code Book with CDMA is dedicated to increase capacity and moreover pointing out the PSNR value. Both methods account for better perception results than referenced method and in one case  $d=2$  the capacity of the Code Book with CDMA is even higher.

The practical application of PN sequences' attributes as autocorrelation and orthogonality were presented in proposed methods. The autocorrelation, in spite of faulty extraction caused by noise in the communication channel, allowed correct identification of PN sequences and properly decomposition of secret message. On the other hand, orthogonality enabled using multiaccess approach in Code Book with CDMA that enhanced the first method in capacity of transferred information through steganographic channel. However, the autocorrelation function of PN sequences used by this method is more

sensitive to faulty extraction bits than Code Book method in steganographic channel.

Applying CDMA technique to steganography system creates highly variable and easily modifying steganography system, which increases security and protection against steganalysis attacks.

#### ACKNOWLEDGMENT

Research described in the paper was financially supported by Ministry of Education of Slovak republic VEGA Grant No. 1/4054/07.

#### REFERENCES

- [1] S. Katzenbeisser and F. A. P. Petitcolas, *Information Hiding Techniques for Steganography and Digital Watermarking*. Norwood: Artech House Publishers, 2000.
- [2] I. J. Cox, M. Miller, J. Bloom, J. Fridrich, and T. Kalker, *Digital Watermarking and Steganography*. 30 Corporate Drive, Suite 400, Burlington, MA 01803, USA: Morgan Kaufmann Publishers is an imprint of Elsevier, 2008.
- [3] S. Dumitrescu, X. Wu, and Z. Wang, "Detection of LSB Steganography via Sample Pair Analysis," *IEEE Transactions on Signal Processing*, vol. 51, no. 7, pp. 408–411, July 2003.
- [4] D. A. Ker, "Steganalysis of LSB Matching in Grayscale Images," *IEEE Signal Processing Letters*, vol. 12, no. 6, June 2005.
- [5] D. Upham, "Jsteg," in *Software available at ftp.funet.fi*, 2000.
- [6] N. Provos, "Defending against statistical Steganalysis," *Proceeding of the 10th USENIX Security Symposium*, pp. 323–335, 2001.
- [7] A. Westfeld, "A Steganographic Algorithm - High Capacity Despite Better Steganalysis, Information Hiding," *Fourth International Workshop, Lecture Notes in Computer Science*, vol. 2137, pp. 289–302, 2001.
- [8] P. Salle, "Model based steganography," 2003, p. 154167.
- [9] A. S. K. Miyake, M. Iwata, "Digital steganography utilizing features of JPEG images," *IEICE Trans. Fundam*, p. 929936, April 2004.
- [10] M. Haque, "A Two-Dimensional Fast Cosine Transform," *IEEE Transactions on Acoustics, Speech and Signal Processing*.
- [11] E. H. Dinan and B. Jabbari, "Spreading codes for direct sequence CDMA and wideband CDMA cellular networks," *Communications Magazine, IEEE*, vol. 36, p. 48–54, September 1998.

# Bi-directional DC/DC converter controlled by microcontroller.

<sup>1</sup>Tomáš BÉREŠ, <sup>2</sup>Ing. Martin OLEJÁR

<sup>1</sup>Dept. of Department of Electrical Drives and Mechatronics, FEI TU of Košice, Slovak Republic

<sup>2</sup>Dept. of Department of Electrical Drives and Mechatronics, FEI TU of Košice, Slovak Republic

<sup>1</sup>beres.tomas@centrum.sk, <sup>2</sup>martin.olejar@tuke.sk

**Abstract** - This thesis describes a proposal of a bi-directional DC/DC converter controlled by Freescale microcontroller. The first part of the thesis is aimed at mathematical model of continuous and discrete PID regulators for DC/DC converters. In the next part is described a created simulation model of DC/DC converter including PSD regulator in Matlab/Simulink environment. Full system is finally tested on physical model outgoing from a development kit DEMOQE.

**Keywords**—Bi-directional DC/DC converter, discrete PSD regulator, Matlab/Simulink, Freescale MCU, MCF51QE128.

## V. INTRODUCTION

The last ten years are characterized by rapid development of microcontrollers and digital signal processors (DSP). It has a great impact on many electronics systems, in which are still used continuous regulators. Especially power electronic is creating a main group of these systems. Using discrete regulators instead of continuous has a lot of benefits, which will by describe later. In this article is describe a basic propose of continuous and discrete regulators used for DC/DC converters.

## VI. BI-DIRECTIONAL DC/DC CONVERTER

A block diagram of bi-directional DC/DC converter controlled by microcontroller is shown in Fig.1. In principle it consists of two basic converters (buck and boost) connected together in one bi-directional DC/DC converter.

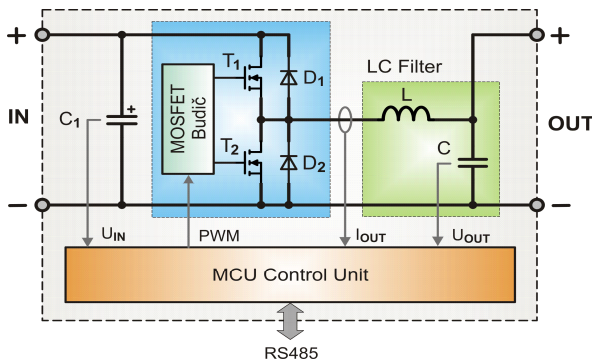


Fig. 1. Block diagram of bi-directional DC/DC converter controlled by microcontroller.

Transistor  $T_1$  and diode  $D_2$  create a buck converter; transistor  $T_2$  and diode  $D_1$  create a boost converter. When transmitting power from  $V_{IN}$  to  $V_{OUT}$  the converter works in buck mode and we can get a lower or equal voltage to input voltage as an output. When transmitting power in opposite direction from  $V_{OUT}$  to  $V_{IN}$  then the converter works in boost mode and we can get a higher voltage as the input voltage as an output.

## III. REGULATORS

The role of the regulator in closed regulation loop is to regulate the system to achieve a divergence between real and requested value as low as possible (ideally equal to zero). While proposing a regulator, the goal is to fulfill this condition not only in the steady state but also in transition state. In the most of the cases regulators of PID (PSD) type are used. With regard to the realization of these regulators we distinguish two main types of regulators:

### A. Continuous

The continuous regulators are realized by operation amplifiers with negative feedback in connection. The basic equation of the PID regulator is as follows:

$$u(t) = K \cdot \left( e(t) + \frac{1}{T_I} \int_0^t e(\tau) d\tau + T_D \frac{de(t)}{dt} \right) \quad (1)$$

$K$  ... is a gain of PID regulator

$T_I$  ... integration constant

$T_D$  ... derivation constant of regulator

$e(t)$  ... regulation offset – difference between required output value from system  $w(t)$  and real output value  $y(t)$ ;  $e(t) = w(t) - y(t)$ ,

$u(t)$  ... output value from regulator

Using Laplace transformation and time response constants  $\varepsilon$ , ( $\varepsilon > 0$ ), transmitting function of PID regulator is:

$$F_R(s) = K \cdot \left( 1 + \frac{1}{T_I s} + \frac{T_D s}{\varepsilon s + 1} \right) \quad (2)$$

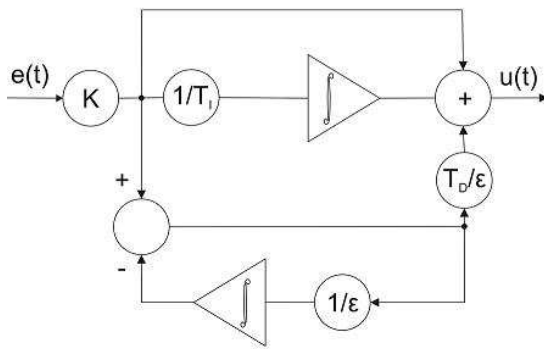


Fig. 2. Diagram of continuous regulator from equation (2)

**B. Discrete**

Discrete PID regulator is realized by software algorithms inside microcontroller. When computing an action intervention for a numerical PID regulator we replace integration by summing **S** and we also replace derivation by difference. The main algorithm of discrete PSD regulator is describes in following equation:

$$u(k) = K \left( e(k) + \frac{T}{T_I} \sum_{i=1}^k e(i) + \frac{T_D}{T} (e(k) - (e(k-1))) \right) \quad (3)$$

An adequate transfer in Z-transformation is:

$$F_R(z) = \frac{a_0 + a_1 z^{-1} + a_2 z^{-2}}{1 - z^{-1}} \quad (4)$$

where  $a_0 = K \left( 1 + \frac{T_D}{T} \right);$

$$a_1 = -K \left( 1 - \frac{T}{T_I} + 2 \frac{T_D}{T} \right); \quad a_2 = K \frac{T_D}{T} \quad (5)$$

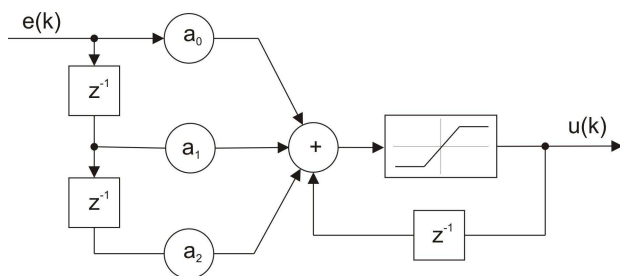


Fig. 3. Block diagram of discrete PSD regulator

**IV. DESIGN AND REALIZATION**

Within the realization of the regulator for the proposed bi-directional DC/DC converter development kit DEMOQE with 32-bit microcontroler MCF51QE128 made by Freescale was used. The main reason for a discrete PSD regulator to be used

is that it allows simple adjustments or full change of control structure with minimal intervention into the hardware part. At this moment the regulator is realised only with voltage control loop as it is shown in Fig.4. By means of the equations (5) and constants which I obtained by the simulation of continuous regulator in Matlab/Simulink I computed the constants  $a_0=25;$   $a_1= -24.6$ . In the future I would like to change control structure to cascade structure where voltage regulation incorporates subordinate current regulation. In this way the limit of maximum current will be provided.

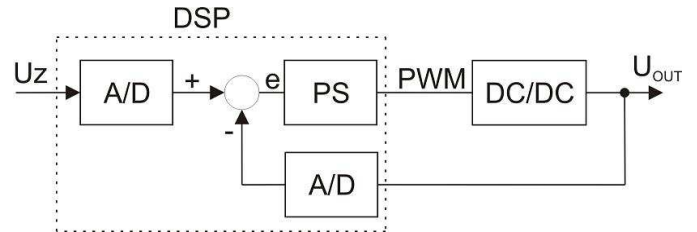


Fig 4. Block diagram of PSD regulator with DC/DC converter.

The activity of voltage regulator was verified in Matlab/Simulink. The block diagram of simulation is shown in the figure 5.

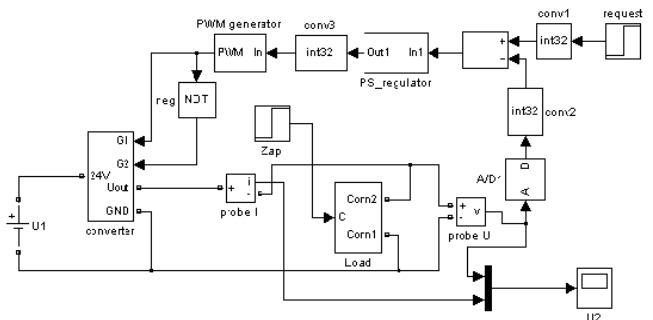


Fig.5 Block diagram of PSD regulator in Matlab/Simulink.

Was applied dynamic change of load on output of converter to demonstrate the activity of the regulator. The simulation results of output voltage and current of converter is shown in the figure 6.

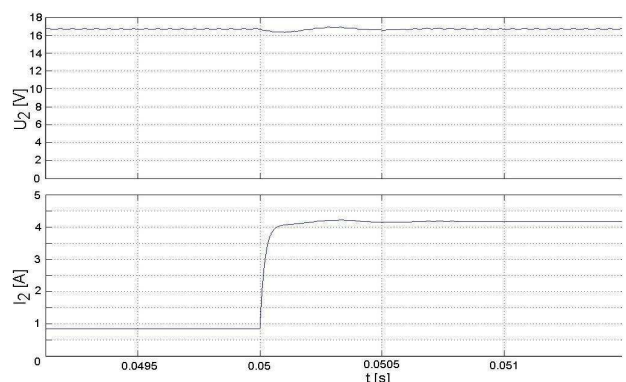


Fig.6 The output of voltage  $U_2$  and current  $I_2$  by applied a dynamic change of the load.

From the voltage behavior is obvious that the regulator successfully regulated an error caused by change of the load with minimal amplitude of the voltage within the period shorter than 1 ms. A comparison of output voltage  $U_2$  with PWM signal on gate of transistor T1 is shown in Fig. 7.

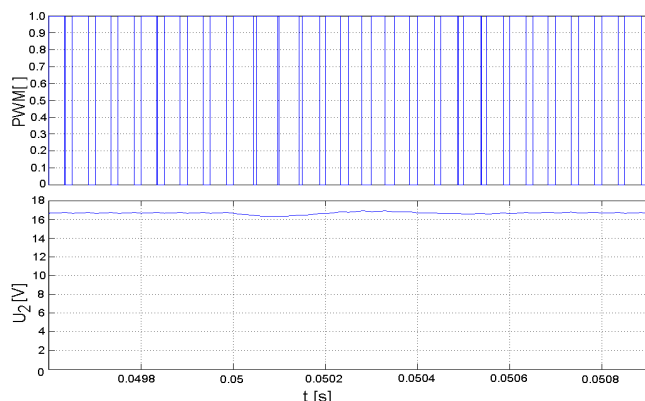


Fig.7 Output voltage  $U_2$  from DC/DC converter and PWM on MCU pin.

The value of output voltage  $U_2$  is direct corresponding with the duty of PWM signal as it is shown in Fig.7. When the voltage on the output of the regulator decreases, the width of PWM impulse increases what prolongs also the closure time of transistor T1.

#### V.CONCLUSION

In this thesis was described a basic design of discrete regulators for power electronic systems like DC/DC converters. In a first part was realized a comparison of discrete regulators with classic continuous regulators. The main focus was oriented on PSD regulator with voltage feedback, which was simulated in Matlab/Simulink and finally tested on physical model outgoing from development kit Freescale DEMOQE. Next research on this area will be oriented more dipper to discrete regulation structures.

#### ACKNOWLEDGMENTS

This work was supported by Slovak Research and Development Agency under project APVV-0095-07 and by Scientific Grant Agency of the Ministry of Education of Slovak Republic under the contract VEGA No. 1/0099/09.

#### REFERENCES

- [1] L.Zboray-F. Ďurovský-J.Tomko, *Regulované pohony*. Viena, 2000 ISBN 80-88922-13-5
- [2] P. Pivoňka, *Číslicová řídicí technika*, Brno, 2002
- [3] B. Smith, "An approach to graphs of linear forms (Unpublished work style)," unpublished.
- [4] E. H. Miller, "A note on reflector arrays (Periodical style—Accepted for publication)," *IEEE Trans. Antennas Propagat.*, to be published.
- [5] J. Wang, "Fundamentals of erbium-doped fiber amplifiers arrays (Periodical style—Submitted for publication)," *IEEE J. Quantum Electron.*, submitted for publication.
- [6] C. J. Kaufman, Rocky Mountain Research Lab., Boulder, CO, private communication, May 1995.
- [7] H. Ertl, J. W. Kolar, F. C. Zach: A novel multi-cell DC-AC converter for applications in renewable energy systems. Norimberg,SRN: PCIM 2001
- [8] Chi K. Tse: Zero-Order Switching Networks and Their Applications to Power Factor Correction in Switching Converters, *IEEE*, 667 – 675 (1997)
- [9] Kácsor, G. - Špánik, P. – Lokšeninec, I.: Simulation Analysis of a Zero Voltage and Zero Current Switching, DC/DC Converter. In proceedings: Transcom '03, Žilina, Slovak Republic, 23 – 25 June 2003, pp. 47-50.
- [10] Marino, P.; Poza, F.; Dominguez, M.A.; Otero, S.: Electronics in Automotive Engineering: A Top-Down Approach for Implementing Industrial Fieldbus Technologies in City Buses and Coaches *IEEE Transactions on Industrial Electronics*, Volume 56, Issue 2, .. 2009, pp. 589 – 600
- [11] Bhattacharya, T.; Giri, V.S.; Mathew, K.; Umanand, L.: Multiphase Bidirectional Flyback Converter Topology for Hybrid Electric Vehicles, *IEEE Transactions on Industrial Electronics*, Volume 56, Issue 1, 2009, pp. 78 – 84
- [12] Emadi, A.; Young Joo Lee; Rajashekara, K.: Power Electronics and Motor Drives in Electric, Hybrid Electric, and Plug-In Hybrid Electric Vehicles, *IEEE Transactions on Industrial Electronics*, Volume 55, Issue 6, 2008, pp. 2237 – 2245
- [13] Todorovic, M.H.; Palma, L.; Enjeti, P.N.: Design of a Wide Input Range DC-DC Converter With a Robust Power Control Scheme Suitable for Fuel Cell Power Conversion, *IEEE Transactions on Industrial Electronics*, Volume 55, Issue 3, 2008, pp. 1247 – 1255
- [14] Gui-Jia Su; Lixin Tang: A Multiphase, Modular, Bidirectional, Triple-Voltage DC-DC Converter for Hybrid and Fuel Cell Vehicle Power Systems, *IEEE Transactions on Power Electronics*, Volume 23, Issue 6, 2008, pp. 3035 – 3046
- [15] Marino, P.; Poza, F.; Dominguez, M.A.; Otero, S.: Electronics in Automotive Engineering: A Top-Down Approach for Implementing Industrial Fieldbus Technologies in City Buses and Coaches, *IEEE Transactions on Industrial Electronics*, Volume 56, Issue 2, 2009, pp. 589 – 600
- [16] Huafeng Xiao; Shaojun Xie: A ZVS Bidirectional DC-DC Converter With Phase-Shift Plus PWM Control Scheme, *IEEE Transactions on Power Electronics*, Volume 23, Issue 2, 2008, pp. 813 - 823



# Introduction to wireless network and Zigbee

<sup>1</sup>Radovan Blichá

<sup>1</sup>Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

<sup>1</sup>radovan.blichá@tuke.sk

**Abstract**— The discovering of revolutionary technologies continues and constant data flow requires increasing demands on computer networks to be non-malfunctioned. Revolution step in development of wireless system is technology ZIGBEE. Zigbee is designed to be reliable, simply implementable and significantly cheaper than other technologies. This paper is focused on introduction of ZigBee.

**Keywords**— Security Standard, Sensor, Sensor Network, ZigBee, Zigbee protocol

## I. INTRODUCTION

The discovering of revolutionary technologies continues and constant data flow requires increasing demands on computer networks to be non-malfunctioned. Science is therefore focused on development of communication protocols for creating of sensorial network consists of a great number of the nodes and to create sensorial node, which could be convenient for this network.

Network built of sensors, which are part of distributed system is called sensor network. Nodes of such sensor network are in most cases randomly allocated in areas of our interest and form so-called sensor field. Sensors in network communicate together using protocols characterized as to be rules for managing syntax, semantics and synchronization of communication. Sensor networks also use layered model like OSI, or TCP/IP networks. Each layer has its own data unit, which is bearer of important data, it processed them and it is also interface for access to higher layer services- Services Access Point (SAP). Communication between layers is handled by a protocol pertaining to given layer.

At the moment, there are several standards such as LAN, Bluetooth, Wi-Fi, etc. but no standard for wireless AD HOC sensor network, which would be simply implementable and adaptable to dynamically changing conditions and would cover specific requirements of the system. AD HOC networks consist of nodes, which communicate each other and can be represented with non-directional graph, in which vertex represents nodes and edge of graph represent connection.

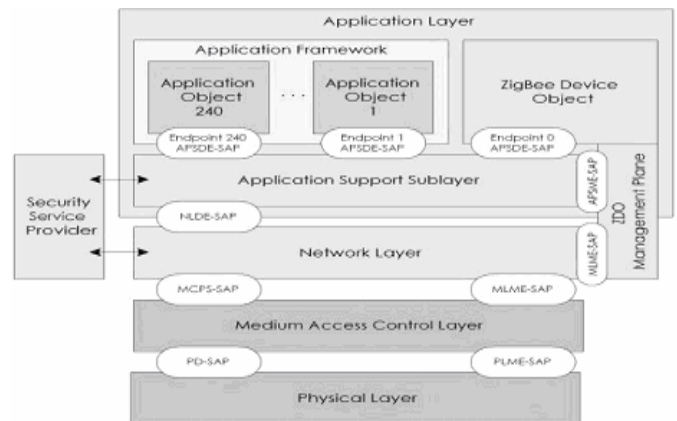


Fig.1 zigbee communication model

Revolution step in development of wireless system is technology ZIGBEE. Zigbee is communication protocol on higher layers for using in radio systems, based on standard IEEE 802.15.4, for wireless network and is especially suitable for transferring large amount of data. Protocol can be used in slow bit rate and low power consumption applications. Zigbee is designed to be reliable, simply implementable and significantly cheaper than other technologies. Zigbee is aimed to be generally used, low cost, self organizing network, which could be used by telecommunication providers, medical facilities, fire and safety alarm systems, systems in intelligent houses, built-in sensors, automobile and army industry. Communication can be encrypted on MAC layer; for example Zigbee device in car is capable to provide secure access to home (remote controlled opening of garage door). Another use is to monitor devices in motion under a closed system - for example identity badges with ZigBee and RFID capabilities mobile within a predefined space.

It takes only several milliseconds to access existing network. That is proof of the flexibility of such network as well. Characteristic sign of Zigbee network is multipath topology. When data can't be received by data terminal equipment through planned path, network dynamically redirects this data to the alternate path.

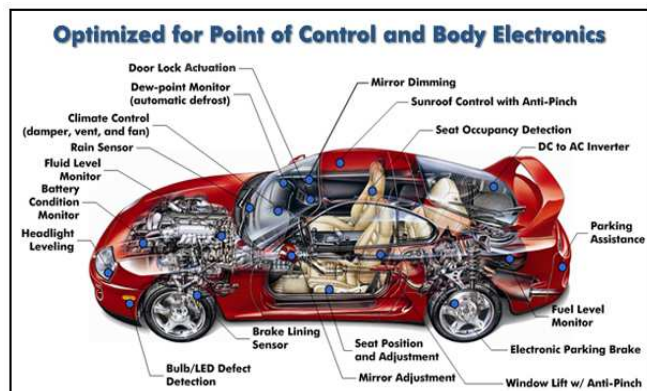


Fig.1 zigbee sensor in the car

Zigbee is a wireless communications protocol that uses lower power and operates in either of three frequencies: 868 MHz (Europe), 915 MHz (USA), 2.4 GHz (mostly worldwide). It's based on an IEEE 802.15.4 standard for WPANs (Wireless Personal Area Networks). There are three types of ZigBee devices: ZC (ZigBee Coordinator), ZR (ZigBee Router), (ZB) ZigBee End Device. Each network has one ZC that maintains network information and security keys. The ZR passes data between devices. The ZED talks to either of the other two types of devices.

Zigbee defines three different network topologies. Basic topology is star topology with a central control node (network coordinator). Second one is tree topology, which allows increasing of distance between coordinator and end point device. Protocol allows creating of redundant connections for topology of mesh type as well. With such topology, it is possible to model network of any shape.

Zigbee network is distinguished by its identifier called PAN-ID. It is 16 bit value randomly assigned during creating of network. Every device is identified on MAC level by its MAC address and on network level by network address. Network has attributes of automatic configuration, self-forming and self-healing. Communication ability represents data capture and sensor nodes controlling, but also nodes are able to cooperate on complex tasks, as monitoring and evaluation of states individual automobiles systems. There is slight latency during transfer of data between two nodes, but from sensor applications point of view, this latency is not so significant.

Zigbee uses basic security standards specified by IEEE (encrypting algorithm AES, CCM mode) and what is more, it defines types of security keys (Master key, Link key, Network key). Security is implemented using 3 layer encryption. Security model checks message integrity on application layer, authentication (so called freshness, or in other words, securing the system against repetitive attacks) and protection against interception. It protects access to devices against attacks from outside (outside the network, where device is active) and inside as well.

Sophisticated solution and design of hardware and software

during creation process of zigbee networks and sensors would be great stimulation for science. Some companies, such as the research, that ZigBee will become the dominant wireless mesh network protocol.

#### REFERENCES

- [1] Cauligi S.; Sivalingam, Krishna M.; Znati, Taieb, "Wireless Sensor Networks," in *Plastics*, 1st ed. 2004. Corr 2nd printing, 2005, 442 p., Hardcover.
- [2] Feng Zhao, and Leonidas J. Guibas (Eds), "Information Processing in Sensor Networks," Proceedings of Second International Workshop, IPSN 2003, Palo Alto, Ca, Usa, April 22-23, 2003.
- [3] Edgar H. Callaway, Jr. and Edgar H. Callaway, "Wireless Sensor Networks: Architectures and Protocols," CRC Press, August 2003, 352 pages.
- [4] R. Bosch GmbH, "Automotive Electrics and Electronics," July 2007, 208-352 pages.
- [5] D.Gislason, "Zigbee Wireless Networking," October 2007

# Direct torque control of asynchronous motor with the help of fuzzy controller

Lubomir CIBULA

Dept. of Electrical, Mechatronic and Industrial Engineering, FEI TU of Košice, Slovak Republic

lubomir.cibula@tuke.sk,

**Abstract**— The present paper deals with designing of a direct torque control of the asynchronous motor by use of a fuzzy controller. The DTC structure has been projected in the Matlab – Simulink program; fuzzy controller has been designed using the Fuzzy Toolbox. Results of the DTC with fuzzy controller are compared with DTC using the Takahashi method.

**Keywords**— direct torque control, fuzzy controller, asynchronous motor

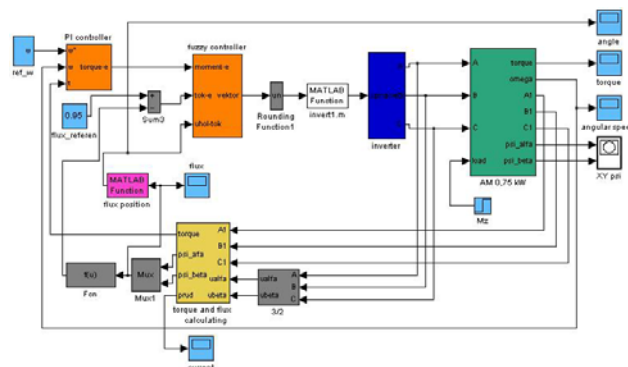


Fig. 1 Structure of the DTC

## I. INTRODUCTION

Direct torque control is one of the most up to date ways of controlling the asynchronous motor. The calculation of necessary stator flux and moment is made based on the measurement of stator voltage and current and with the help of a block of the calculation of magnetic flux and moment. The objective of this work is to achieve better results than in case of DTC that was realized according to method of Takahashi. The improvement of the results relates mainly to the increase of regulation's precision and decrease of undesired great vibration of moment.

## II. THE DIRECT TORQUE MOMENT STRUCTURE

The entire DTC structure consists of several principal blocks:

- The asynchronous motor model block (termed as AM 0,75 kW)
- Block of the indirect frequency converter with voltage inter-circuit supplied by which is the AM (inverter)
- The magnetic flux and torque calculation block (torque and flux calculating)
- Angle (position) of the stator magnetic flux calculation block (flux position)
- Block of the PI speed controller (PI controller) and of the fuzzy torque and flux controller (fuzzy controller)

## III. FUZZY CONTROLLER

### A. Competence functions

Fuzzy controller (FC) was designed and implemented using the Fuzzy Logic Toolbox program (Matlab). Used as input values for the FC were: torque error (Fig. 2.), error of the stator flux (Fig. 3.) and the stator flux position (Fig. 4.).

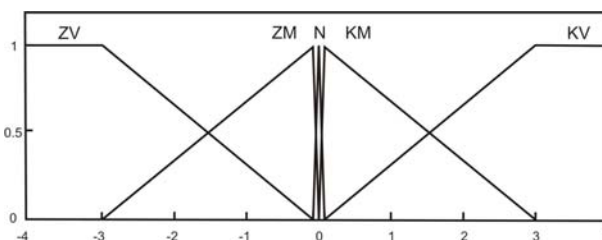


Fig. 2 Distribution of competence function for error of the moment

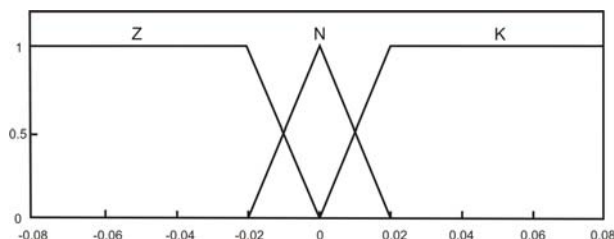


Fig. 3 Distribution of competence function for error of stator flux

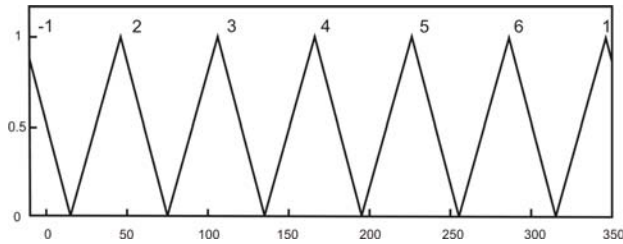


Fig. 4 Distribution of competence function for position of stator flux

The fuzzy controller output is the required stator voltage vector (Fig. 5). The resultant vector (which is obtained from the fuzzy controller) does not present directly required vector (it is not an integer), and hence it must be further adjusted (rounded) within the Rounding function block. This final value already presents the stator voltage vector that is needed as an input into the indirect frequency converter with voltage inter-circuit.

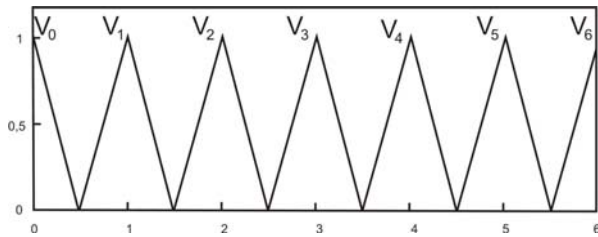


Fig. 5 Distribution of the competence function for the vector of stator voltage

**B. Production rules**

The very operation of the fuzzy system – controller, is based on extrapolation (inference) rules, similarly as it is the case with expert systems. The advantage of this presentation of the knowledge is transparency and easy “readability”. These rules are of IF-THEN type and are generated in the Rule editor (one of the tools of the Fuzzy Logic Toolbox program). Overall number of the rules is 105 and are illustrated with the help of Look-up table (Fig. 6)

-1(1)					2					3				
Fe	Te	P	ZE	N	Fe	Te	P	ZE	N	Fe	Te	P	ZE	N
PL	1	2	2		PL	2	3	3		PL	3	4	4	
PS	2	2	3		PS	3	3	4		PS	4	4	5	
ZE	1	0	4		ZE	2	0	5		ZE	3	0	6	
NS	6	5	4		NS	1	6	5		NS	2	1	6	
NL	6	5	5		NL	1	6	6		NL	2	1	1	
4					5					6				
Fe	Te	P	ZE	N	Fe	Te	P	ZE	N	Fe	Te	P	ZE	N
PL	4	5	5		PL	5	6	1		PL	6	1	1	
PS	5	5	6		PS	6	6	1		PS	1	1	2	
ZE	4	0	1		ZE	5	0	2		ZE	6	0	3	
NS	3	2	1		NS	4	3	2		NS	5	4	3	
NL	3	2	2		NL	4	3	3		NL	5	4	4	

Fig. 6 Look-up table with production rules

**IV. COMPARISON OF DTC METHOD SIMULATIONS (TAKAHASHI) WITH DTC REALIZED WITH FUZZY CONTROLLER**

The waveforms were simulated for the start up of AM.

In case of this simulation we start up the AM to nominal speed 145 rad/s. We start up the motor from zero speed. The waveform of angular speed is shown in Fig. 7. Due to the fact that during the waveform of angular velocities for individual methods there are no significant differences among the methods there is illustrated only the waveform of angular speed in DTC with the help of fuzzy.

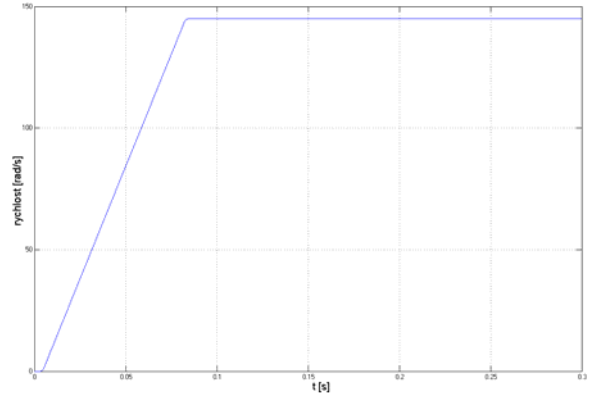


Fig. 7 Angular speed waveform of DTC method with the help of fuzzy

More noticeable differences between the methods are represented by the moment waveform. On Fig. 8 is the comparison of the waveform of moment at Takahashi method of DTC and at method of DTC with the help of fuzzy. We can see from the waveform that the moment at Takahashi method is noticeably more oscillated as in case of DTC method with the help of fuzzy. This is one of the main advantages of this method in comparison with Takahashi method.

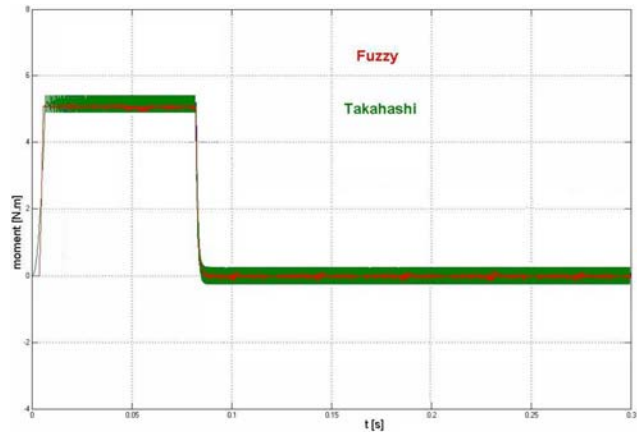


Fig. 8 Comparison waveforms of the moment for individual methods

In respect to the fact that in case of DTC with the help of fuzzy we try to maintain the constant magnetic flux at the required moment, it is obvious also the waveform of the vector of magnetic flux in coordinates X-Y on Fig. 9.

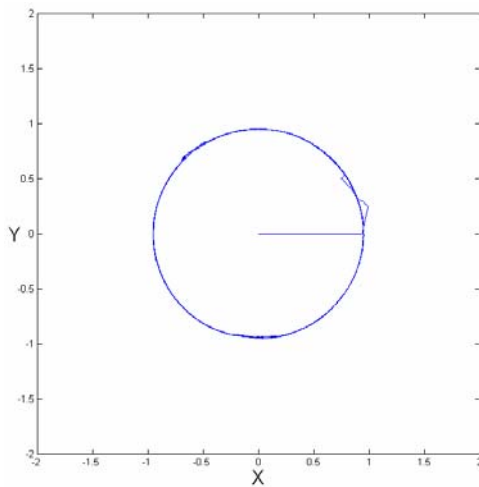


Fig. 9 Waveform of the vector of magnetic flux in coordinates X-Y for method with the help of fuzzy

Magnetic flux is shown also in the following figure. Magnetic flux is shown in time base and we can see one in his components  $\alpha$ ,  $\beta$ . (Fig. 10). As we can see in fig. 10, waveforms of magnetic fluxes are smoothed, which results also from previous waveforms (Fig. 9).

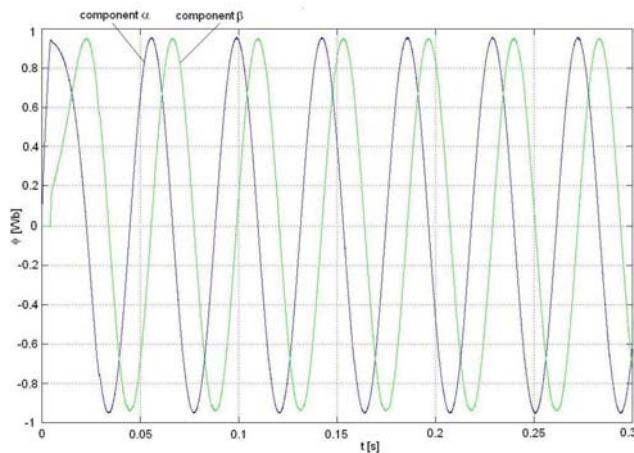


Fig. 10 Waveform of magnetic flux in components  $\alpha$ ,  $\beta$

In the next figure waveform of stator current in components  $\alpha$ ,  $\beta$  is shown (Fig. 11). From figure we can see the near smoothed harmonic waveform and current peak at start up of AM.

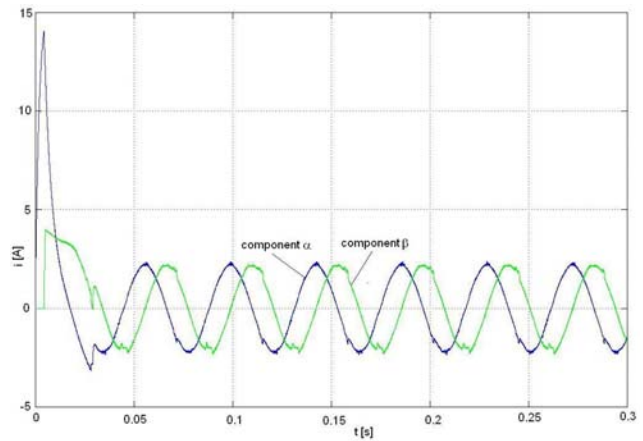


Fig. 11 Waveform of stator current in components  $\alpha$ ,  $\beta$

Parameters of motor: Siemens 1LA7083 – 4AA10

$$U = 400 \text{ V}$$

$$I = 1.93 \text{ A}$$

$$P = 0.75 \text{ kW}$$

$$R_1 = 12.5 \text{ } \Omega$$

$$R_2' = 10.583 \text{ } \Omega$$

$$L_1\sigma = L_2\sigma = 0.0201 \text{ H}$$

$$L_1 = L_2 = 0.4322 \text{ H}$$

$$L_h = 0.412085 \text{ H}$$

$$J = 0.0017 \text{ kgm}^2$$

$$n_n = 1395 \text{ rpm}$$

$$\cos \varphi = 0.8$$

## V. CONCLUSION

This article deals with direct torque control of asynchronous motor with the help of fuzzy controller and its comparison with Takahashi method of DTC. The comparison of Takahashi method with DTC with the help of fuzzy is based on their waveforms. The waveforms were simulated for the start up of asynchronous motor.

In waveforms of Takahashi's method of DTC obvious is the noticeable oscillation of the moment in comparison with fuzzy method. This high oscillation causes in operating conditions great noise and also excessive vibration not only of motor's shaft but also contingent vibration of the whole system. As far the required angular speed is concerned, it can be achieved with these methods in approximately the same time and individual waveforms do not differ very much from each other.

## ACKNOWLEDGMENT

The assignment has been solved within the grant VEGA No. 1/4076/07

## REFERENCES

- [1] J. Timko, J. Žilková, P. Girovský, "Electrical drives," ISBN 80-8073-529-8, Košice.

- [2] M. Žalman, I. Kuric, "Direct torque and flux control of induction machine and fuzzy controller," *Journal of Electrical Engineering – No.1. Volume 5.2005.*
- [3] M. Depenbrock, "Direct Self Control (DSC) of Inverter-Fed Induction Machines," *IEEE Trans. Power Electronics, Vol. PE-3, No.4, pp 420-429, 1988.*
- [4] P. Brandštetter, L. Štěpanec, "Fuzzy Logic Control of Induction Motor Drive," *IWCIT'01, VŠB-TU Ostrava, 2001, ISBN 80-7078-907-7.*
- [5] Timko J., Žilková J., Girovský P.: *Electrical drives. TU Košice, ISBN 80-8073-529-8, pp 149, Košice 2007.*
- [6] Timko J., Žilková J., Girovský P.: *Shaf sensorless vector control of an induction motor. In: Acta Technica ČSAV. Vol.52, no.1 (2007), p.81-91. ISSN 0001-7043*
- [7] Vittek, J. et al.: *Comparison of sliding mode and forced dynamics control of electric drive with flexible coupling employing PSMM. In: IEEE International Conference on Industrial Technology IEEE ICIT 2008: 21-24 April 2008. Piscataway: IEEE 2008, p.1-6. ISBN 978-1-4244-1706-3.*
- [8] Žilková J., Timko J., Girovský P.: *Nonlinear system control using neural networks. In: Acta Polytechnica Hungarica. Vol.3, no.4 (2006), p.85-94. ISSN 1785-8860*

# Micro and Macro Analysis of Polarization Current.

<sup>1</sup>Vieroslava ČAČKOVÁ, <sup>2</sup>Lýdia DEDINSKÁ, <sup>3</sup>Milan KVAKOVSKÝ

<sup>1</sup> Dept. of Electric Power Engineering, FEI TU of Košice, Slovak Republic

<sup>2</sup> Dept. of Electric Power Engineering, FEI TU of Košice, Slovak Republic

<sup>3</sup> Dept. of Electric Power Engineering, FEI TU of Košice, Slovak Republic

<sup>1</sup>Vieroslava.Cackova@tuke.sk, <sup>2</sup>Lydia.Dedinska@tuke.sk, <sup>3</sup>Milan.Kvakovsky@tuke.sk

**Abstract**— This article deals with the dielectric properties of insulation, degradation processes, the influence of thermal aging and evaluating the state of isolation through dependences of charge and discharge characteristics.

**Keywords**— Degradation, thermal aging, charge and discharge characteristics

## I. INTRODUCTION

The electrical equipment in service is steered a gradual degradation of its insulating properties. Degradation is non-reversible change in the functional characteristics of the materials and components as a result of in service operations, and leads ultimately to the situation where the component cannot fulfill its task anymore and will fail. This phenomenon is called aging and has a significant role in the life of the entire electrical system. Aging processes is a summary of the physical and chemical changes occurring in service. Status isolation substantially affects the remaining life management of the system, therefore necessary is to diagnose what stage in its life is the device. Among the diagnosis methods are belonged direct-current measuring methods.

## II. DEGRADATION MECHANISM

The degradation mechanism can be classified in mechanical, electrical and thermal stresses, which is relation to time, a comprehensive exposure and conditions of the insulation exposed.

### A. Mechanical stress

The mechanical stress is divided as follows:

- centrifugal forces
- mechanical vibration,
- electrodynamic forces.

Electrodynamic forces have permanent character in the current load, but there are occurred also sudden changes in the electrical circuit. Current and voltage pulses induce them.

### B. Electrical stress

The insulation of equipment is exposed to the Electrical stresses on the entire period of service. Increasing electric field intensity decreases life management of insulation depending on the time. In the classification of the impact of electrical stress on the equipment it is necessary to distinguish the impact of DC and AC electric field.

Insulating material is divided into two basic groups according the influence of electric field on the material:

- Thermoreaktive
- Thermoplastic

*Thermoreaktive* are characterized by a threshold value, below which the life of material is theoretically infinite, if there is another kind of stress.

*Thermoplastic* - there is observed any threshold value, below is not occurred aging. Stabilization life curve was not found even in very long time. It is assumed that in these materials is a threshold value of electrical stress too. However is it so small that is generated only by the extreme long time to go beyond the technical use of the material.

### C. Thermal stress

Key aspects are included the sequence of thermal aging of chemical and physical changes as a sequence of chemical degradation reactions. The process towards thermodynamic equilibrium. Temperature is a factor with essential significance for electrical equipment service capability. There is affected exposure to high or low ambient temperatures, elevated temperatures above ambient temperature, and the increase of dielectric losses arising in the isolation of electrical equipment. So thermal insulation overload causes the accelerated aging and reduce its Electro-insulating properties.

In solid insulations is led to irreversible changes.

The time that elapses before limiting value when the insulation be unable to perform safely the required function can be written by Arrhéni [1]:

$$\tau = A \exp \frac{B}{T} = \tau_0 \left( \frac{E}{E_0} \right)^{-n} \quad (1)$$

where:  $\tau$  - the life time,

$\tau_0$  - the reference life time,

T - absolute temperature,

A, B - constants are determined activation energy of reaction,

E - electric stress,

$E_0$  - threshold electric stress,

n - constant of voltage resistance contiguous to thermal.

Relationship of Arrhéni is based on the dependence to the speed of chemical reactions from first-order temperature.

Montsinger rule is applied to most materials and equipment. It says that increasing the temperature about 8 K to 10 K shortens the lifetime of the isolation system by half.

#### D. Combined stress

An experimental research was found that the multiple degradation processes for the material at the same time, is much greater the degradation effect as the sum of the particularly individual load. For combined stress true [2]:

$$\frac{\tau}{\tau_0} = \frac{\tau_t}{\tau_0} \cdot \frac{\tau_e}{\tau_0} \Rightarrow$$

$$\tau = \tau_0 \left( \frac{E}{E_0} \right)^{-n} \exp(-B\Delta T) \quad (2)$$

where:  $\tau_t$  – the thermal life time,  
 $\tau_e$  – the electric life time.

### III. DIAGNOSTIC METHODS

Diagnostic methods for determining the properties of insulating materials during service can be divided according to the effect on measured object [3]:

- non-destructive methods
- potentially destructive methods
- destructive methods

DC electrical measuring methods are classified analysis the dielectric response are among the potentially destructive methods. There are includes curves of charge and discharge currents and method of recovery of tension.

Systems in real environment are operated to replace the Mathematics- Physics models and insulation system can be replaced by its equivalent model. Maxwell-Wagner model dielectric consists of a system resistors and capacitors.

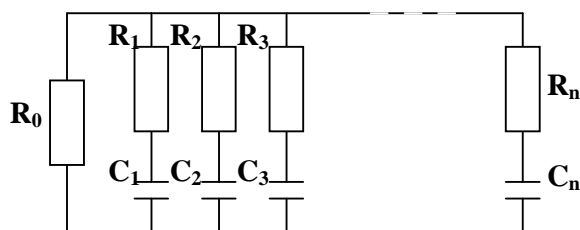


Fig. 1. Equivalent diagram of dielectric [4]

This model is described the behavior of the system in field of time, respectively frequency, and enable its physical application. On the basis this model and knowledge of the polarization phenomena in the structure of the material dielectrics was developed, isothermic -Relaxation Current Analysis (IRC-Analysis), which can be used to evaluate the change in electro physical structure material.

#### A. Microscopic view of the IRC analysis

Each process aging is operated on material thereby change the microstructure and composition of insulators. This is resulted from the characteristic changes of behavior in dielectric relaxation.

In the IRC analysis is observed an increasing amount slower polarization processes during aging.

In the transport mechanisms of current is occurred to "generation" of the trapping levels through the operation of an electric field on a constant temperature in dielectrics. Consequently is come into existence to return of the

breakdown levels monotonically decreasing current. Electric field transmits direct charges of one ion and electron mold and has been directed in this direction dipoles on the corresponding relaxation period until a new equilibrium

In the polymer are located catch levels over total zone holes as well as are distributed discrete energy levels for the isothermal relaxation current. Disregard charge carrier of recombination can be the direct determination catch levels of the measured relaxation current. Activation energy catch levels  $W_T$  is determined by the relationship [5]:

$$W_T = W_L - W_H \Rightarrow$$

$$W_T(t) = W_L - W_H = k \cdot T \cdot \ln(v \cdot t) \quad (3)$$

where:  $W_L$  - energy of lower conductivity band,  
 $W_H$  - energy catch levels,  
 $k$  - Boltzmann constant,  
 $v$  - thermal electron velocity.

Than can be written for initialization occupation of catch levels and energy spectrum of catch levels:

$$\Gamma(W) \cdot V(W) = \frac{2}{q \cdot d \cdot k \cdot T} \cdot t \cdot \left[ I_0 + \sum_{i=1}^3 a_i \cdot e^{-\frac{t}{t_i}} \right] \quad (4)$$

where:  $\Gamma$  - initialization occupation of catch levels,  
 $V$  - energy spectrum of catch levels,  
 $q$  - charge,  
 $d$  - thickness of isolation,  
 $a$  - constant of material.

#### B. Microscopic view of the IRC analysis

A substantial part of polarization range of material is in the field  $10^{-3} - 10^5$  s. This area can be explored DC methods based on monitoring the voltage and current responses.

#### MEASUREMENT CHARGE AND DISCHARGE CURRENTS

Dielectric on the DC voltage connects to flowing current, whose size depends on [6]:

- size of the enclosed voltage
- time of connection voltage
- state of isolation

Electrical conductivity and polarization processes influence passing current. Conduction processes are characterized leakage currents. Polarization processes are characterized to flowing absorption current in dielectric.

Line current in dielectric is considered as the sum of elementary relaxation currents exponentially decreasing in time.

For the passing charging current  $i_n(t)$  in dielectric is valid after connecting direct voltage [7]:

$$i_n(t) = i_c(t) + i_a(t) + i_v(t) \quad (5)$$

where:  $i_n(t)$  - charging current,  
 $i_c(t)$  - capacitance current,  
 $i_a(t)$  - the absorption current (polarization),  
 $i_v(t)$  - leakage currents (conductivity).

Typical curved line of the current response to the DC voltage connection is in figure 2



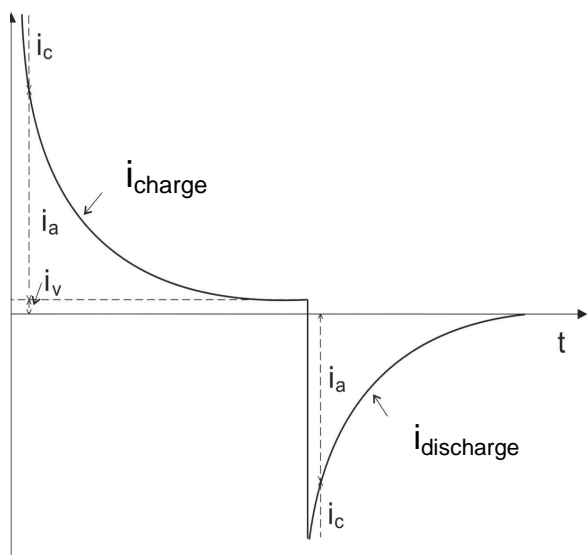


Fig.2 Current response to an induced pulse [2]

• *capacitance current*  $i_c(t)$  - flows of geometric capacitance  $C_0$  after the charging time, its value decreases from the moment impress voltage.

*absorbent stream*  $i_a(t)$  - is connected with the transport fixed charge between non-homogeneous segments of insulation, which is decreasing with time,

*leakage currents*  $i_v(t)$  resp. conductivity component of stream - this stream flows through the insulating resistance  $R_0$ . It is connected with the existence of free charge carriers in dielectrics. The current fluctuation is at the beginning in consequence to voltage changes on the insulation before is reached steady state and later takes a constant value  $I_v$ .

After the voltage disconnection leads to the discharge of capacity composition of material and macroscopic makes itself felt the discharging current, which consists of resorbing  $i_r(t)$  and capacity  $i_c(t)$  current [1]:

$$i_v(t) = i_r(t) + i_c(t) \quad (6)$$

Line current in dielectric is considered as the sum of elementary relaxation currents exponentially decreasing in time. Of Figure 1 can be written:

$$i(t) = \frac{U}{R_0} + \sum_{i=1}^n I_{mi} \exp\left(\frac{-t}{\tau_i}\right) \quad (7)$$

where:  $U$ - applied direct potential,  
 $R_0$ - resistance by direct potential,  
 $I_{mi}$  - amplitude of  $i$ - element Debyes elementary current,  
 $\tau_i$  - the time of relaxation constant of  $i$ - element Debyes elementary current.

#### IV. CONCLUSION

IRC analysis method is allowed to calculate the elements of the replacement model dielectric on the basis of parameters defined the number of exponential components that are obtained from the analysis. The replacement dielectric model is made possible to calculate the dielectric loss factor respectively comprehensive permittivity in frequency domain where are two additional parameters, which can determine the material degradation.

The method can be applied to various kinds of solid and liquid insulators. However, it is necessary to obtain the

characteristic processes in the aging process and apply them to specific insulating material. Application of this method is trouble free in the industry.

#### ACKNOWLEDGMENT

This work was supported by scientific Grant Agency of the ministry of Education of the Slovak Republic project VEGA No. 1/0368/09 and APVV-20-006005.

#### REFERENCES

- [1] CIMBALA, R.: Starnutie vysokonapäťových izolačných systémov, FEI-TU, EQUILIBRIA, Košice 2007
- [2] CIMBALA, R.: Starnutie izolačných systémov vysokonapäťových strojov, Anotácia monografie, FEI-TU Košice
- [3] [www.tuke.sk/fei-kee/predmety/DvEE/Diag2.doc](http://www.tuke.sk/fei-kee/predmety/DvEE/Diag2.doc)
- [4] BARTÁK, A., MRAVINÁČ, L., NEUMAN, J.: Diagnostika poruch izolací elektrických strojů, STNL, PRAHA, 1984
- [5] LELÁK, J.: Starnutie, stanovenie zvyškovej životnosti a kvalifikácia elektrotechnických prvkov a zariadení, STU, Bratislava, 1996
- [6] MARTON, K.: Diagnostika VN a VVN zariadení, ČSVTS, Stará Lesná, 1987
- [7] KOLCUNOVÁ, I.: Diagnostika v elektroenergetike, Prednášky pre 5. Roč. EE/VREE, KEE Košice, 2006, Dostupné na internete: <http://web.tuke.sk/fei-kee/predmety/dvee.html>

# Dependence of Dissipation Factor on Voltage of Liquid Dielectrics

<sup>1</sup>Lýdia DEDINSKÁ, <sup>2</sup>Vieroslava ČAČKOVÁ, <sup>3</sup>Milan KVAKOVSKÝ

<sup>1</sup>Dept. of Electric Power Engineering, FEI TU of Košice, Slovak Republic

<sup>2</sup>Dept. of Electric Power Engineering, FEI TU of Košice, Slovak Republic

<sup>3</sup>Dept. of Electric Power Engineering, FEI TU of Košice, Slovak Republic

<sup>1</sup>Lydia.Dedinska@tuke.sk, <sup>2</sup>Vieroslava.Cackova@tuke.sk, <sup>3</sup>Milan.Kvakovsky@tuke.sk

**Abstract**—The aim of this work was to investigate the electrical properties of vegetable oils and to compare them with the most commonly used insulating liquids. The dissipation factors ( $\text{tg } \delta$ ) of different types of oils were compared. The dissipation factor was measured at increased voltage (0,1 – 2 kV) and changed in temperature (20 – 100 °C).

**Keywords**—dissipation factor, electrical properties, liquid dielectrics, vegetable oils

## I. INTRODUCTION

The majority of high voltage transformers are filled with liquids that work as an electrical insulators as well as a heat transfer media. Insulation system of electrical power transformer is based on mineral oil now. The popularity of mineral transformer oil is due to availability and low cost, as well as being an excellent dielectric and cooling medium. Ever since the world oil reserves were tapped in the 1940s, petroleum products have become widely available.[1] Petroleum-based products are so vital in today's world that we cannot imagine a time we may not have them easily available. Transformers and other oil-filled electrical equipment use only a tiny fraction on the total petroleum consumption, yet even this fraction is almost irreplaceable. But there is a problem with liquidation of these transformer oils. This is the reason why we should alternatively use other insulating oils. Their ecological point is another reason to try to find other substitute liquids. For that purpose, lots of activities have been initiated. This work is oriented to the investigation of the vegetable oils[1].

## II. DISSIPATION FACTOR $\text{tg } \delta$

We can imagine an isolation of high-voltage devices as dielectric of condenser which electrodes are made as high-voltage and low-voltage part of device. By connecting of the alternating voltage to the condenser, the current will start to flow inside of dielectric. Perfect dielectric is not showing any losses and vector of charging current has capacitance nature only, phased for 90° from connected voltage vector.

But insulating materials used in electrical devices are not perfect. Current flowing via dielectric consists both from

reactive component and power component, which is created by conductivity and polarization of dielectric.

Power component of current is caused by following effects:

- resistivity of insulating material has large resistivity, but not infinite.
- dielectric becomes polarized inside electrical field.

Polarizing current flowing via dielectric consists both from reactive component and power component, see figure 1. [2]

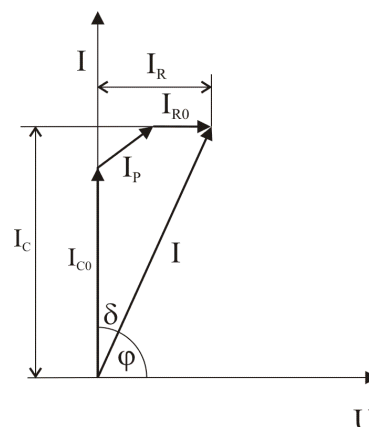


Fig. 1. Phasor diagram of real condenser. [2]

Coefficient of dissipation factor  $\text{tg } \delta$  is defined as tangent of the angle where differs phase displacement between real dielectric and perfect dielectric. The value of dielectric losses depends on dielectric losses and that is why we can consider this value as tool for measuring the quality of insulation system. According to phasor diagram (fig. 1): [2]

$$\text{tg } \delta = \frac{I_R}{I_C} \quad (1)$$

### A. Dependence of dissipation factor on voltage

Increasing of voltage connected to insulation will cause change of some processes, which have influence on the value of dielectric losses. We are able to observe these phenomena mostly in case when above specific voltage level, the different mechanism of losses will start which is caused by ionizing processes. Level of voltage where the curve  $\text{tg } \delta = f(U)$  is bending itself up is marked as ionizing threshold or the offset voltage of discharge  $U_{\text{poc}}$ . [2]

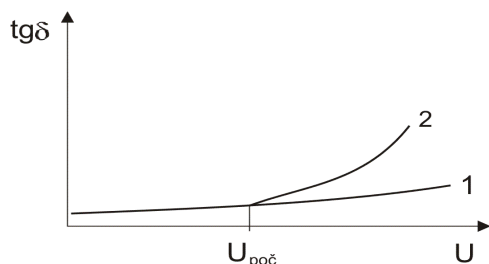


Fig. 2. dissipation factor as a function of. [2]

By measuring of dependency  $tg \delta = f(U)$  in periodical intervals we are able to consider changes which occurred during the machine operation[2]

### III. PREPARATION OF SPECIMENS AND MEASUREMENTS

#### A. Measurement dissipation factor

Dissipation factor  $tg \delta$  was measured by Schering fully automatically bridge by TETTEX AG type 2818 with the frequency of 50Hz. Samples were placed in electrode system Tettex 2903/AT, temperature was controlled by Tettex 2965/ZH from 20°C to 100°C. The HV bridge has had a guaranteed sensitivity of dissipation factor better than  $\pm 2 \cdot 10^{-4}$ . The applied voltage was changed from 0.1kV up to 2.0kV by step 0.2kV. First measurement was measured at a room temperature. After that the temperature was increased up to 100°C by step 10°C. For each temperature level was made the voltage dependence of dissipation factor.

#### B. Samples

Investigations were focused on the measurement of the dissipation factor of vegetable oils and comparison to mineral based transformer oil and silicon oil. As vegetable oil were used two types of fluids: sunflower oil and colza oil. As transformer oil was used inhibit oil ITO100. As silicon oil was used LUKOSIOL M 200. The following samples were measured:

- M1 – inhibit transformer oil,
- S1 - non-filtrated sunflower oil,
- R1 - colza oil Raciol
- SiL1 - silicon oil LUKOSIOL M 200

TABLE I  
MEASURED VALUES OF THE DISSIPATION FACTOR  $TG \delta$  FOR EACH KINDS OF SAMPLES

U [ kV ]	tan $\delta$ [ - ]			
	M1	S1	R1	SiL1
0,2	0,0049	2,06	0,037	0,0012
0,4	0,0037	2,09	0,0363	0,0008
0,6	0,0028	2,13	0,0366	0,0001
0,8	0,0026	2,15	0,0376	0,0001
1	0,0026	2,18	0,0383	0,0001
1,2	0,0026	2,19	0,0391	0,0001
1,4	0,0026	2,22	0,0406	0,0001
1,6	0,0026	2,25	0,042	0,0001
1,8	0,0026	2,27	0,0432	0,0001
2	0,0027	2,3	0,044	0,0002

### IV. MEASURING RESULTS

For different types of oils and the repeated measurements the voltage dependence loss factor  $tg \delta$  was built. These characteristics of different types of oil were compared among themselves.

#### A. Mineral oil

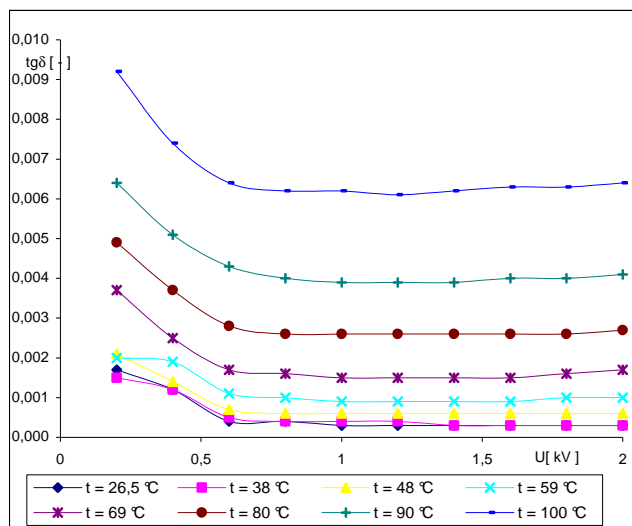


Fig. 3. Dissipation factor as a function of voltage at temperature 20°C for the M1

From fig.3 for mineral oil ITO 100 is visible, that value of  $tg \delta$  drops down to voltage of 0,6kV for all temperatures. By further increasing of voltage the value of  $tg \delta$  is constant.

The dissipation factor  $tg \delta$  rises with the temperature. Up to temperature of 69C is the value of dissipation factor is very close and varies between 0.001 – 0.002. Step changes occur from a temperature of 80 ° C to 100 ° C. At 100 ° C the value is 0.009.

#### B. Sunflower oil

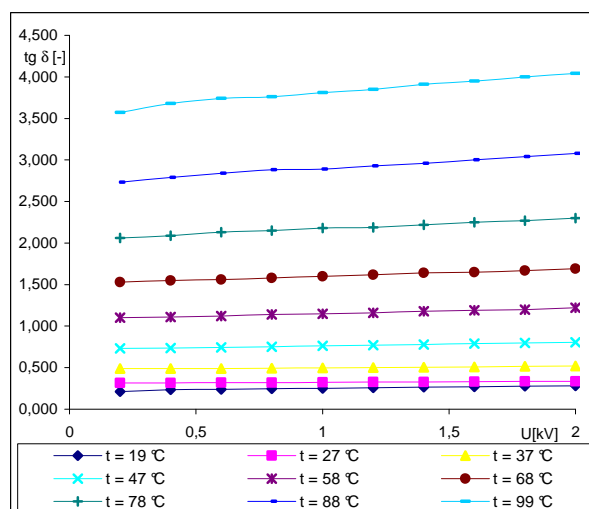


Fig. 4. Dissipation factor as a function of voltage at temperature 20°C for the S1

The Figure 4 depicts the dependency for the measurement of sunflower oil, S1. Values of dissipation factor depending on the voltage for the temperatures up to 68 ° C can be considered as constant. Significant changes occurred at higher temperatures - the value of  $tg \delta$  is raising together with rising

voltage. The dissipation factor for a temperature of 78 °C was 2.0. Raising the temperature by ten degrees it has increased to 3.0 and by further increase of 10 degrees at a voltage 2 kV, it reached value of 4.

C. Raciol oil

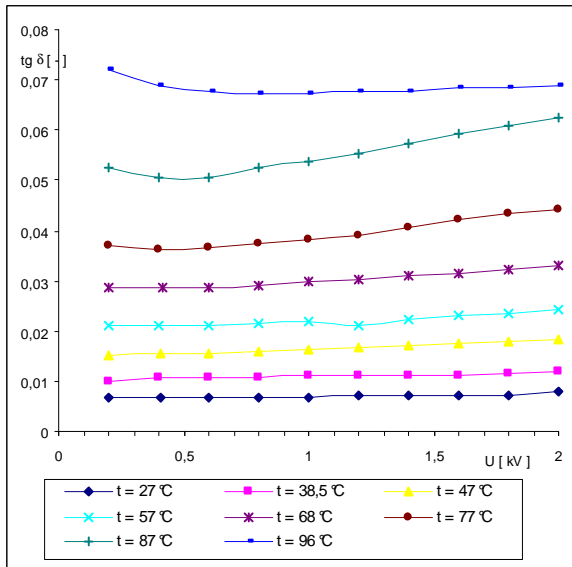


Fig. 5. Dissipation factor as a function of voltage at temperature 20°C for the R1

Regarding the sample of colza oil R1, for temperatures up to 60°C in dependency of voltage the  $tg \delta$  has constant values. Step changes of dissipation factor occur over the temperature of 87°C where its value is 0.06 at voltage of 2kV.

D. Silicon oil

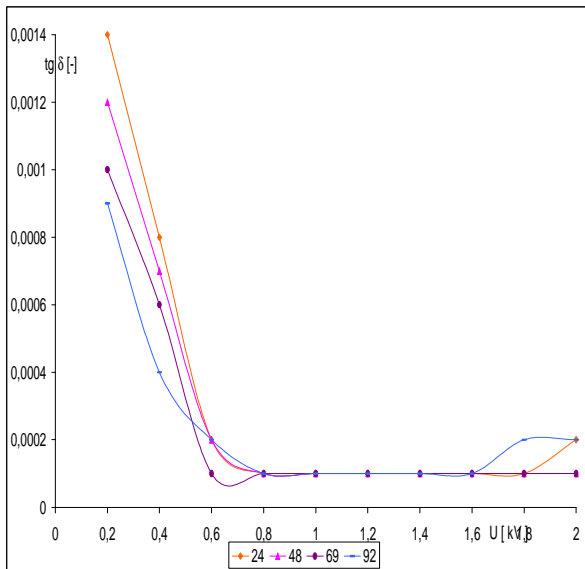


Fig. 6. Dissipation factor as a function of voltage at temperature 20°C for the SiL

For silicon oil the value of dissipation factor grows with the temperature (at voltage of 0.2kV). Up to 0.6kV it is value declines and by further increasing of voltage it is value remains constant at 0.0001.

E. Comparison of different types of oil samples

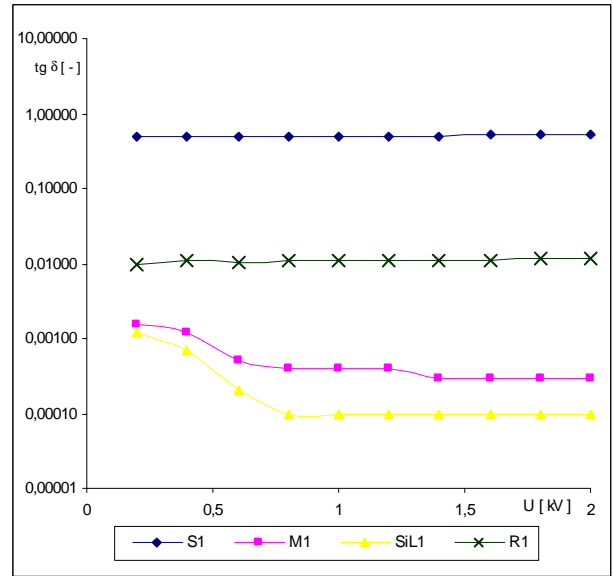


Fig. 7. Dissipation factor as a function of voltage at temperature of 40 °C for different oil samples

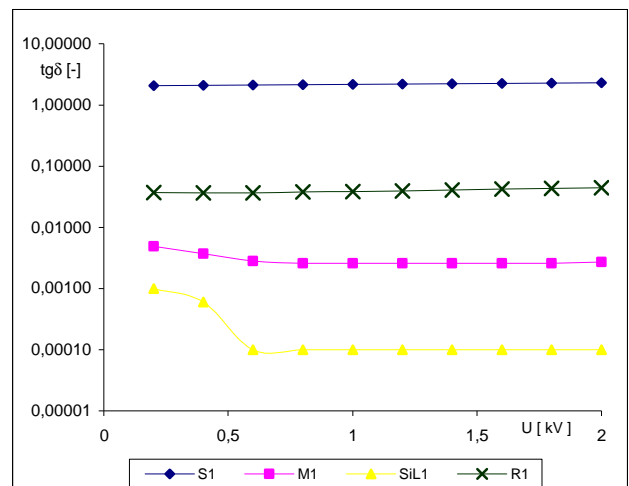


Fig. 8. Dissipation factor as a function of voltage at temperature for different oil samples at 80 °C

The lowest value of dissipation factor has silicon oil LUKOSIOL M 200 ( $tg \delta=0.0001$ ). For mineral oil the value is 0.001. Regarding the vegetable oils, the value of dissipation factor  $tg \delta$  is:

- for colza oil  $tg \delta=0.006$
- for sunflower oil  $tg \delta=0.23$

V. CONCLUSION

The purpose of this work was to compare voltage dependency of dissipation factor for liquid dielectrics. As natural ester fluids were used two types of fluids: sunflower oil and colza oil. Samples were achieved directly from manufacturer. But it is necessary to add that natural ester fluids samples were without any additives. Natural oils were compared with Transformer oil ITO 100 and silicon oil LUKOSIOL M 200.

Measurement results of dissipation factor  $\text{tg } \delta$  are at enclosed charts. Measurements were performed on 4 samples. From charts is visible, that the value of dissipation factor  $\text{tg } \delta$  of liquid insulators rises with temperature. It was also shown, that  $\text{tg } \delta$  rises only very slowly with the value of applying voltage. By lower values of voltage up to 0.6kV for transformer oil ITO 100 (M1) and silicon oil (SiL1) the dissipation factor decreases. But for higher values of voltage it remains constant. For vegetable oils, the dissipation factor remains constant up to 80°C. Over the 80°C it rises. The values for colza oil are comparable to placed values of transformer oil.

It is needed to make more measurements to say that vegetable oils are ready to be used in practice. There are more important properties to study for example breakdown voltage, electrical stability, permittivity.

#### ACKNOWLEDGMENT

This work was supported by scientific Grant Agency of the ministry of Education of the Slovak Republic project VEGA No. 1/0368/09 and APVV-20-006005.

#### REFERENCES

- [1] Oommen, T.V.: Vegetable Oils for Liquid-filled Transformers. In.: Magazine Electrical Insulation IEEE, volume 18, number 1, january/February 2002, pp.6-11
- [2] Kolcunová, Iriada: Diagnostika elektrických strojov, Technická univerzita Košice, 2006, ISBN 80-8073-550-.
- [3] Poljak, František: Dielektriká izolanty, ALFA, 1983
- [4] Tichý, V. - Pallo, V. - Mašek, V.: Transformátorový olej. , STNL, Bratislava, 1962 .
- [5] Cimbala,R., Kršňák,I, Kolcunová,I.: The Computation of Infulence of Steady Element on Polarization Spektrum, Journal Acta Polytechnica Prague, Vol. 43, No. 2/2003, ISSN 1210-2709, Prague

# Microstructure development of SnAgCu solder joint

Juraj ĎURIŠIN

Department of Technologies in Electronics, FEI TU of Košice, Slovak Republic

juraj.durisin@tuke.sk

**Abstract**—Microstructure of SnAgCu solder alloy results from composition of the alloy and solidification process. In case of solder joint is very important not only composition of the solder alloy, but also interaction between the alloy and soldered surface. Another important factor is effect of ageing process, causing significant changes in microstructure of the joint. Investigation of the microstructure evolution results in better knowledge of the joint behaviour in various use conditions.

**Keywords**— ageing process, intermetallic compounds, microstructure, solder alloy.

## I. INTRODUCTION

Microstructure of SnAgCu (SAC) solder alloys results from weight (atomic) content of constituent elements (or atoms), properties of the elements and solidification process. SAC alloy is a solid solution of metals with limited solubility.

Very important is not only composition of solder alloy, but also interaction (caused by diffusion) between the alloy and soldered surface. Soldered surface (pad on printed circuit board, PCB) is a copper foil coated with various coatings – surface finishes. Typical composition of surface finishes is: Ni/Au, Sn, solder alloy, Ag, etc. The interaction has fundamental impact on properties of the final solder joint. Microstructure of the final solder joint results from constituent elements of solder alloy and pad, time of soldering and solidification process. Solder joint (interaction solder/pad) is solution of metals with limited solubility.

Limited solubility results in formation of intermetallic compounds (IMCs), significantly influencing mechanical properties and lifetime of solder joints.

In our experiments we focused on microstructure of solder volume (and solder/pad interface) after soldering and after ageing process.

## II. LIMITED SOLUBILITY

Limited solubility of two metals results from properties of the alloy constituent atoms (Hume-Rothery rules) [1]:

- The crystal structures of solute and solvent must match.
- The atomic radius of the solute and solvent atoms must differ by no more than 15%.
- The solute and solvent should have similar electronegativity. If the electronegativity difference is too great, the metals will tend to form intermetallic compounds instead of solid solutions.
- Maximum solubility occurs when the solvent and solute have the same valence. Metals with lower valence will tend to dissolve metals with higher valence, vice versa the metals will tend to form intermetallic compounds

instead of solid solutions.

These rules are summarized in I.

TABLE I  
PROPERTIES OF SAC SOLDER CONSTITUENT ATOMS [2]

Atom	$\beta$ -Sn	Ag	Cu	Ni
Crystal system	tetragonal	cubic	cubic	cubic
Atomic radius [pm]	145	160	135	135
Electronegativity	1.96	1.93	1.90	1.91
Valence	4	1	2	2

Evaluation of I leads to II.

TABLE II  
EVALUATION OF MUTUAL COMBINATION OF SAC SOLDER CONSTITUENT ATOMS

Elements combination	$\beta$ -Sn - Ag	$\beta$ -Sn - Cu	$\beta$ -Sn - Ni
Crystal system	×	×	×
Atomic radius [pm]	•	•	•
Electronegativity	•	•	•
Valence	×	×	×

From these evaluation (II) result binary phase diagrams, in our case we are interested in SnAg, SnCu and SnNi diagrams. Other combinations of elements (solder/pad) form mainly solutions with unlimited solubility in full range of their concentrations.

As it is clear from the phase diagrams (Fig. 1 to 3), there are formed IMCs in wide range of concentrations. It means, that in soldering are the IMCs essential part of solder (or solder/pad) microstructure.

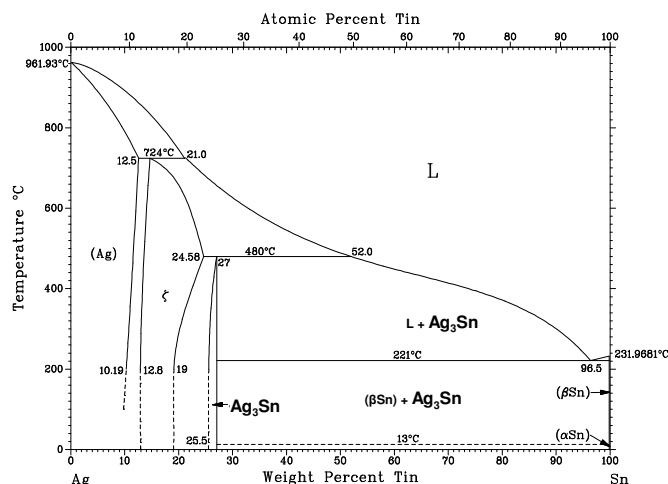


Fig. 1. SnAg binary phase diagram [3].

In SnAg diagram (Fig. 1) is clearly present  $\text{Ag}_3\text{Sn}$

intermetallic (orthorhombic crystal system). Ag<sub>3</sub>Sn is in solder volume located in inter-dendritic area. SnAg diagram is eutectic type of diagram.

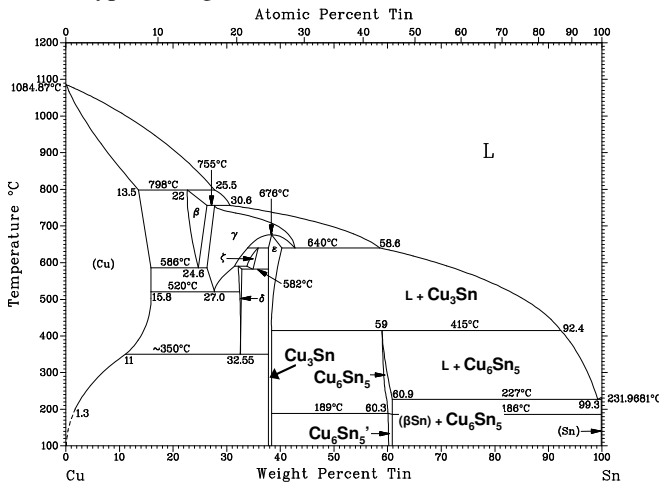


Fig. 2. SnCu binary phase diagram [3].

In SnCu diagram (Fig. 2) are clearly present Cu<sub>3</sub>Sn (orthorhombic) and Cu<sub>6</sub>Sn<sub>5</sub> (hexagonal crystal system) intermetallics. Cu<sub>6</sub>Sn<sub>5</sub> is located in solder volume and at the solder/Cu pad interface, Cu<sub>3</sub>Sn only at the solder/Cu pad interface. SnCu diagram is eutectic type of diagram.

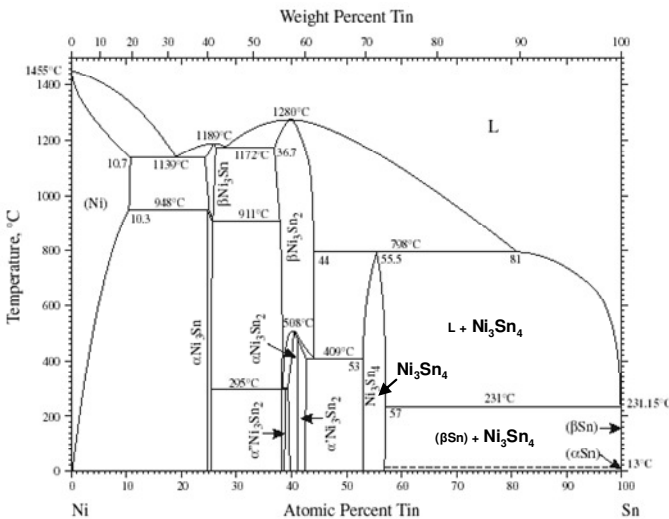


Fig. 3. SnNi binary phase diagram [3].

In SnNi diagram (Fig. 3) is clearly present Ni<sub>3</sub>Sn<sub>4</sub> intermetallic (monoclinic crystal system). Ni<sub>3</sub>Sn<sub>4</sub> is located at the solder/Ni pad interface. SnNi diagram is peritectic type of diagram.

TABLE III

PHYSICAL PROPERTIES OF SOLDER JOINT CONSTITUENT METALS [4]

Property	Sn	Ag	Cu	Ni
<b>Mechanical character</b>	ductile	ductile	ductile	ductile
<b>Thermal conductivity [W.m<sup>-1</sup>.K<sup>-1</sup>]</b>	66.8	429	401	90.9
<b>Resistivity [10<sup>-8</sup> Ω.m]</b>	11	1.587	1.72	6.99
<b>Density [g.cm<sup>-3</sup>]</b>	7.365	10.49	8.96	8.908
<b>Thermal expansion [μm.m<sup>-1</sup>.K<sup>-1</sup>]</b>	22.0	18.9	16.5	13.4

TABLE IV

PHYSICAL PROPERTIES OF INTERMETALLIC COMPOUNDS [4]

Property	Cu <sub>6</sub> Sn <sub>5</sub>	Cu <sub>3</sub> Sn	Ni <sub>3</sub> Sn <sub>4</sub>	Ag <sub>3</sub> Sn
<b>Mechanical character</b>	brittle	brittle	brittle	brittle
<b>Thermal conductivity [W.m<sup>-1</sup>.K<sup>-1</sup>]</b>	34.1	70.4	19.6	-
<b>Resistivity [10<sup>-8</sup> Ω.m]</b>	17.5	8.93	28.5	-
<b>Density [g.cm<sup>-3</sup>]</b>	8.3	8.9	8.65	-
<b>Thermal expansion [μm.m<sup>-1</sup>.K<sup>-1</sup>]</b>	16.3	19.0	13.7	-

Intermetallic compounds significantly influence properties of solder joints. Properties of the IMCs usually fundamentally differ from properties of constituent metals (III, IV).

Summary of properties of the IMCs:

- a) mechanically brittle,
- b) low (lower) thermal conductivity (compared to the constituent metals),
- c) low (lower) electrical conductivity (compared to the constituent metals).

### III. EXPERIMENTAL PROCEDURE AND RESULTS

Reflow soldering was realized in vapour phase reflow soldering machine. The applied peak temperature was 230°C, time of soldering was 240 seconds.

As a solder material was used Alpha Metal solder paste containing 96.5Sn3Ag0.5Cu alloy. PCB pad was coated with HASL (Hot Air Solder Leveling) and chemical Ni/Au surface finish. During soldering Au dissolves into solder, therefore Au is not considered later. The selected samples were after the soldering thermally aged 1000 hours at the temperature of 125°C. Scratch patterns of samples were observed by light and electron microscopy. Phase identification was performed by energy-dispersive X-ray spectroscopy (ratio of elements).

SAC solder/pad interface is depicted in the next figures. It is clearly visible difference between HASL and Ni/Au surface finish (after soldering). In case of HASL copper diffuses into solder (Fig. 4), forming Cu<sub>6</sub>Sn<sub>5</sub> layer at the interface and elongated Cu<sub>6</sub>Sn<sub>5</sub> in volume of solder. Thick Ni layer (Fig. 5) is well observed at the copper foil. Between solder and Ni layer is formed Ni<sub>3</sub>Sn<sub>4</sub> intermetallic.

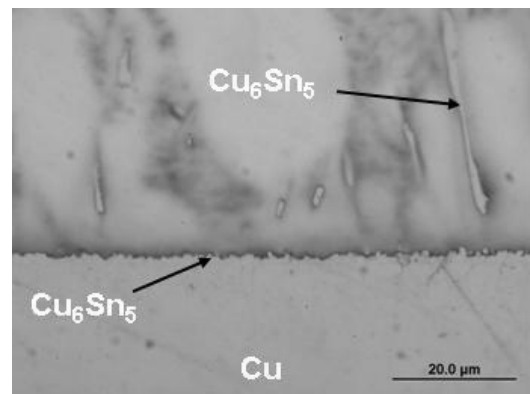


Fig. 4. SAC solder/Cu pad interface.

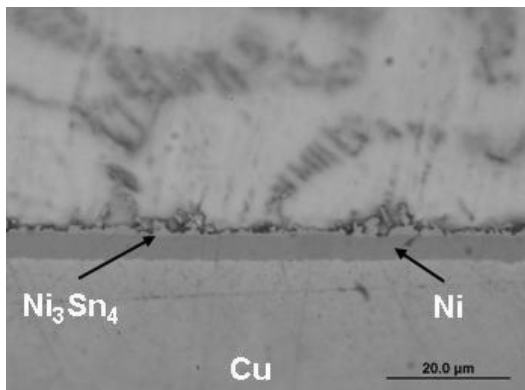


Fig. 5. SAC solder/Ni pad interface.

After the ageing (125°C, 1000 hours) the intermetallic layer changed significantly only in case of Cu pad (Fig. 6).  $\text{Cu}_6\text{Sn}_5$  layer has considerably grown, additionally between  $\text{Cu}_6\text{Sn}_5$  layer and Cu foil was formed thin  $\text{Cu}_3\text{Sn}$  layer.  $\text{Ni}_3\text{Sn}_4$  intermetallic retained the same thickness (Fig. 7).

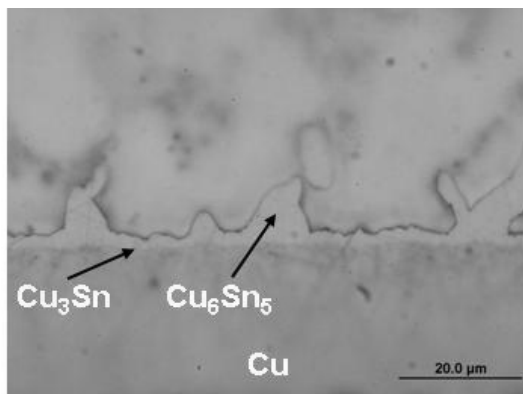


Fig. 6. SAC solder/Cu pad interface.

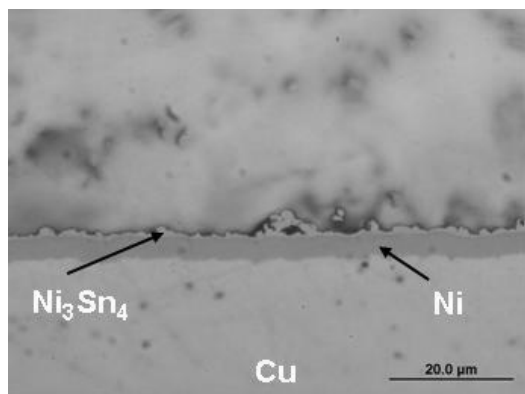


Fig. 7. SAC solder/Ni pad interface.

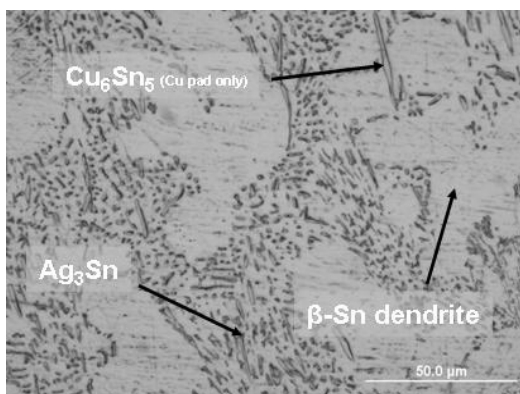


Fig. 8. Microstructure of SAC solder volume, after soldering.

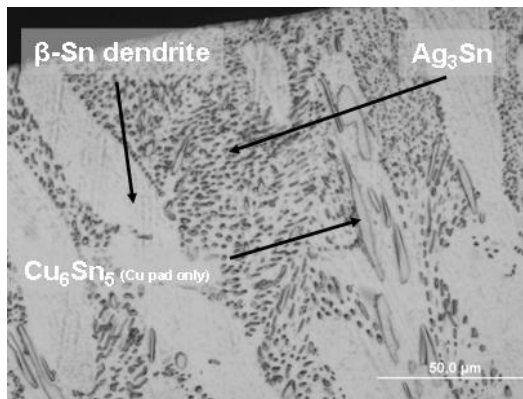


Fig. 9. Microstructure of SAC solder volume, after the ageing.

In the solder volume (Fig. 8) there were not such significant changes in microstructure compared to microstructure after the ageing (Fig. 9), although some coarsening of IMCs is observed. In inter-dendritic area is present  $\text{Ag}_3\text{Sn}$  and  $\text{Cu}_6\text{Sn}_5$  intermetallic. If soldered on bare copper (HASL), elongated  $\text{Cu}_6\text{Sn}_5$  can be found also in volume of  $\beta\text{-Sn}$  dendrites.

#### IV. CONCLUSION

The realized experiments and following analysis confirmed assumes, that microstructure of lead-free solder joints strongly depends on composition of solder alloy, composition of PCB pad and the ageing process. Significant change was observed if compared Ni/Au and HASL (bare copper) surface finish. The diffusion rate of Cu into molten (and solid) solder is much higher compared to Ni. It was exactly proved by analysis of volume of solder (effect of soldering - extensive presence of elongated  $\text{Cu}_6\text{Sn}_5$ ) and solder/pad interface (effect of ageing - extensive presence of  $\text{Cu}_6\text{Sn}_5$  intermetallic layer).

#### ACKNOWLEDGMENT

This work has been supported by foundation of project VEGA 1/0298/09.KEGA, SK-CZ 0065-07 and KEGA 3/6465/08.

#### REFERENCES

- [1] "Hume-Rothery rules". [http://en.wikipedia.org/wiki/Hume-Rothery\\_rules](http://en.wikipedia.org/wiki/Hume-Rothery_rules).
- [2] "Periodic table". [http://en.wikipedia.org/wiki/Periodic\\_Table](http://en.wikipedia.org/wiki/Periodic_Table).
- [3] "Phase Diagrams & Computational Thermodynamics". <http://www.metallurgy.nist.gov/phase/solder/solder.html>
- [4] R. J. Fields, S. R. Low, "Physical and mechanical properties of intermetallic compounds commonly found in solder joints", Metallurgy Division, NIST. [http://www.metallurgy.nist.gov/mechanical\\_properties/solder\\_paper.html](http://www.metallurgy.nist.gov/mechanical_properties/solder_paper.html)



# Architecture for Traffic Sign Detection

<sup>1</sup>Martin FIFIK,

<sup>1</sup>Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

<sup>1</sup>martin.fifik@tuke.sk

**Abstract**—This paper describe architecture for traffic sign detection for use in moving cars. Input data are converted in HSV or RGB color space. In each color space are done preprocessing actions. This gives us several of ROI, where can traffic sign be. From every ROI features are extracted. This is done by projection transforms. After all this steps, the class of traffic sign is choose. Potential of this proposed detection system is discussed.

**Keywords**—traffic sign, projection transformations, image classification.

## I. INTRODUCTION

There are several possible options how to create new architecture of traffic sign recognition system [1-3,7]. Systems that we generally consider can use:

- Advantages of the GPS systems,
- A low price traffic sign information receiver in a car,
- Visual information and explore it.

GPS systems in nowadays is common equipment of cars. Big advantage of this system is that it can locate position of the objects very precisely. But what is missing is a database and map of updated traffic sign on the roads. Create this database may be difficult.

Another option is use a receiver in the car, which is able to receive information from road traffic signs. The issue that arises here is problem with installing low price transmitters on every traffic sign. This method may be expensive, because the numbers of traffic signs are enormous.

A last option that we considered in our development is use and exploring visual information. This system will require a camera and processor with additional software and hardware installed in the car. Advantage of this system is in easy modification.

To acquisition of a traffic scene we can use color cameras or gray-scale cameras. For gray-scale camera we can use only shape detection method, but when we use color camera, we can use color segmentation and shape detection. Generally, also the system build on the color camera will be faster

To process visual information, two steps detection scheme is used (Fig. 1).

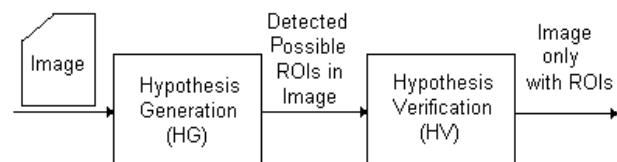


Fig. 1. Two steps detection scheme

Traffic sign detection is very seriously problem in nowadays. All car producers are trying to develop and implement traffic sign detection system in their cars.

The rules for safety traffic are displayed on traffic signs. Traffic signs are designed to show us some rule or warn us before something. If we leave out some traffic sign, we can put us in danger situation or worst; we can be participant in car accident. An automatic road sign detection system will be helpful, that can warn us when we leave out some traffic sign [2, 3, 7].

Traffic sign follow strictly shape and color. For that, can be good recognizing from surrounding environment, while driving.

## II. ARCHITECTURE OF THE PROPOSED SYSTEM

Based on this two steps scheme, was designed architecture of traffic sign recognition (Fig. 2).

This system can be divided in two subsystems. One, which is called fast, composed from CSR block and Sign Class Classification block, and second slower, composed from Image Segmentation, Invariant Feature Extraction System, Feature Memory, Feature Modification and Sign Type Classification blocks.

Fast subsystem is used to promptly detection of class type and position of traffic sign. Slower subsystem is used to completely recognition of traffic sign type.

### A. Function

Surrounding traffic scene is recorded with camera. In our case color camera is used. From camera created data stream, key frame is extracted. Key frame is used in next processing.

This key frame is processed in CSR block, where the base features from image are extracted. Features extracted from key frame are color, shape and ROIs position. This preprocessing is very fast, because no special transformation is there needed. As reference color space we use RGB color space. By extracting color, CSR block generates 3 binary maps. Three

maps because we follow three colors: ed, blue and yellow. In this bitmaps we looking for shapes of traffic signs. If traffic sign was detected, then from color and shape features traffic sign class type is identified. Traffic class type goes together with position of traffic sign to block Image Segmentation. From key frame the Region Of Interest (ROI) can be segmented. This makes an input to block called Invariant Feature Extraction System.

In Invariant Feature Extraction System are used projecting transform, like Trace transform, Hough transform [4-6, 8]. Invariant features are features that have equal values in cases when the image is rotated, moved, resized or have some other similar modification. The block of Invariant Feature Extraction System with cooperation with blocks Feature Memory and Feature Modification block, gives information for Sign Type Classification block. In this block are brought information about class type of detected traffic sign. An output of this block gives us completely recognized traffic sign for block called Traffic Sign Interpretation.

### III. EXPERIMENTS

The classification was done by Euclid classification (1D and 2D). Number of tested images was 320. The overall recognition rate depends on choosing functional in Trace transform process. In tests was used 9 different group of Trace functional. For moved images average success recognition rate was 90,14%.

### IV. CONCLUSION

This paper propose architecture of invariant system for traffic signs recognition with mediate requirements on computational power. This system contains two subsystems, faster and slower. Faster is designed to help prevent driver before immediate danger, a slower is designed to give driver completely information about traffic sign. So this system can be used as warning system before danger or as system for fully recognizing traffic signs.

### ACKNOWLEDGMENT

This work was partially supported from the grants VEGA No. 1/0034/09, project COST IC0802.

### REFERENCES

- [1] A. Laika, W. Stechele, *A review of different object recognition methods for the application in driver assistance systems*, Image Analysis for Multimedia Interactive Services, 2007, WIAMIS, 6-8 June 2007.
- [2] F. P. Paulo, L. P. Correia, *Automatic detection and classification of traffic signs*, Image Analysis for Multimedia Interactive Services, 2007, WIAMIS, 6-8 June 2007.
- [3] A. Broggi, P. Cerri, P. Medici, P. P. Porta, *Real Time Road Signs Recognition*, Intelligent Vehicles Symposium, 2007 IEEE Volume , Issue , 13-15, 981 – 986, June 2007.
- [4] J. Turán, D. Šišková, J. Turán, Jr., P. Filo, L. Ovseník, *Trace Transform and KLT Based Invariant Features and Image Recognition System*, Acta Electrotechnica et Informatica, No.3, Vol.6, 5-15, 2006.
- [5] J. Turán, *Fast translation Invariant Transforms and their Application*, Elfá, Košice, 1999.
- [6] M. Nixon, A. Aguado *Feature extraction and image processing*, Academic Press, 2008.
- [7] A. Kadyrov, M. Petrou, *The Trace Transform and Its Applications*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.23, No.8,811-828, August 2001.
- [8] N. Otsu. *A threshold selection method from gray-level histogram*, IEEE Transactions on Pattern Analysis and machine Intelligence, vol. 25, n°8,

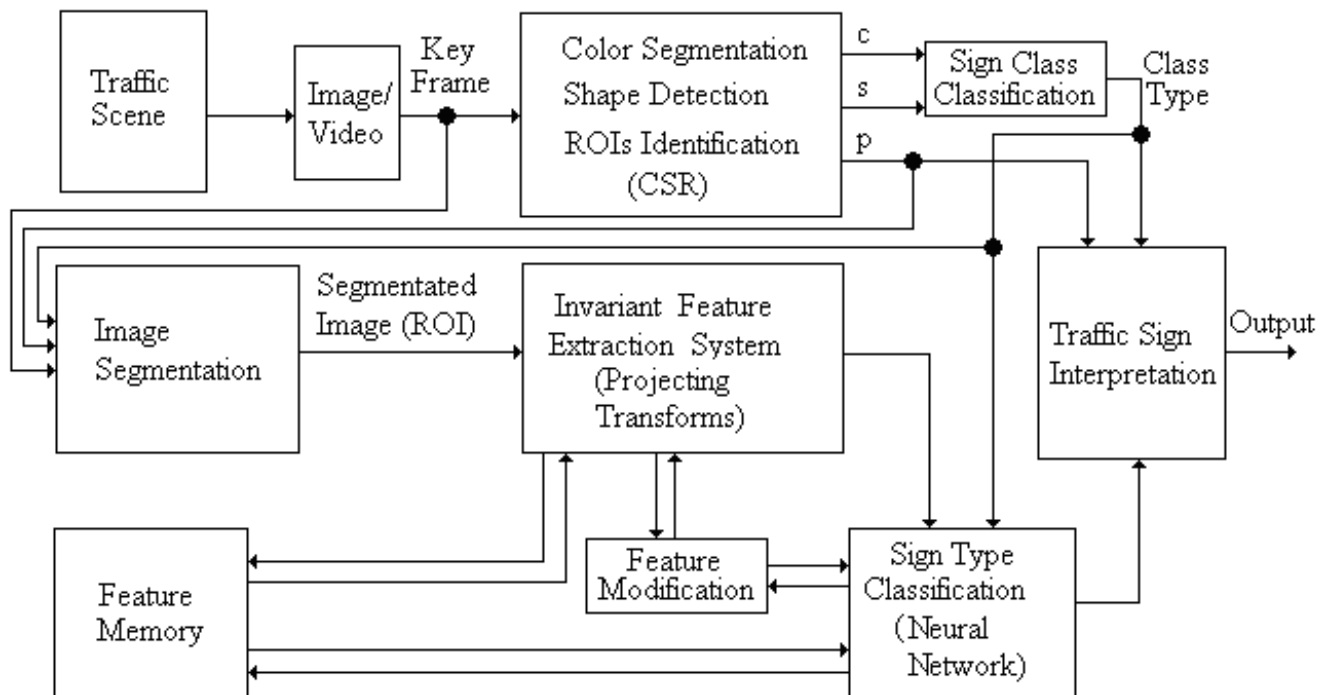


Fig 2. Proposed System for Traffic Sign Recognition



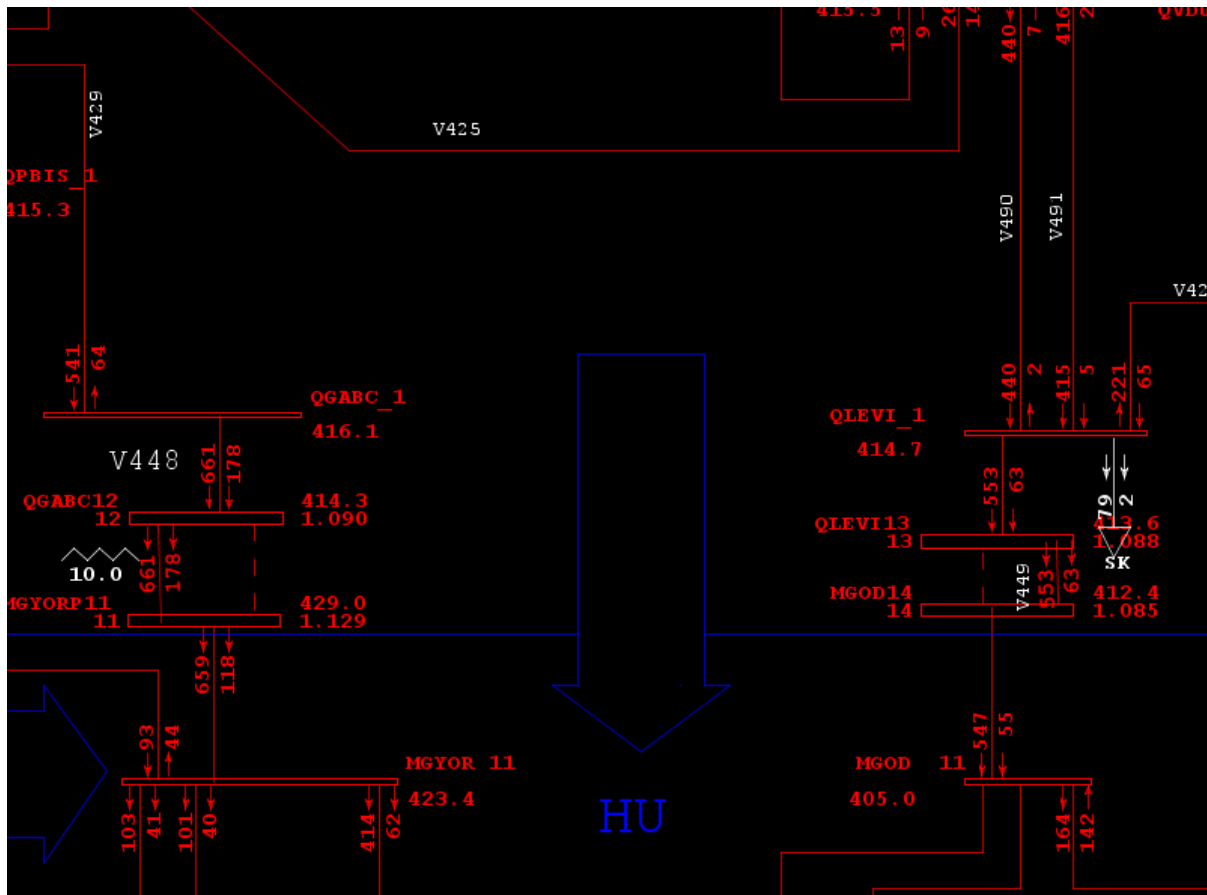


Fig. 3 Flows of active power as a function of PST in line V448

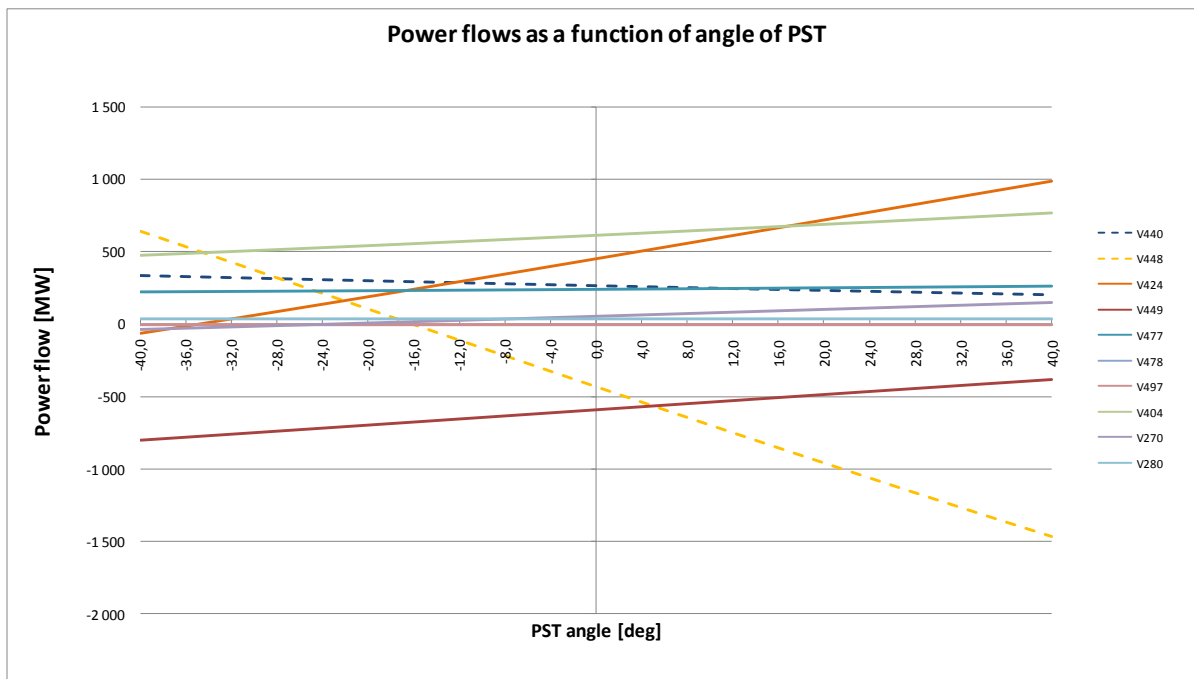


Fig. 4 Flows of active power as a function of PST in line V448



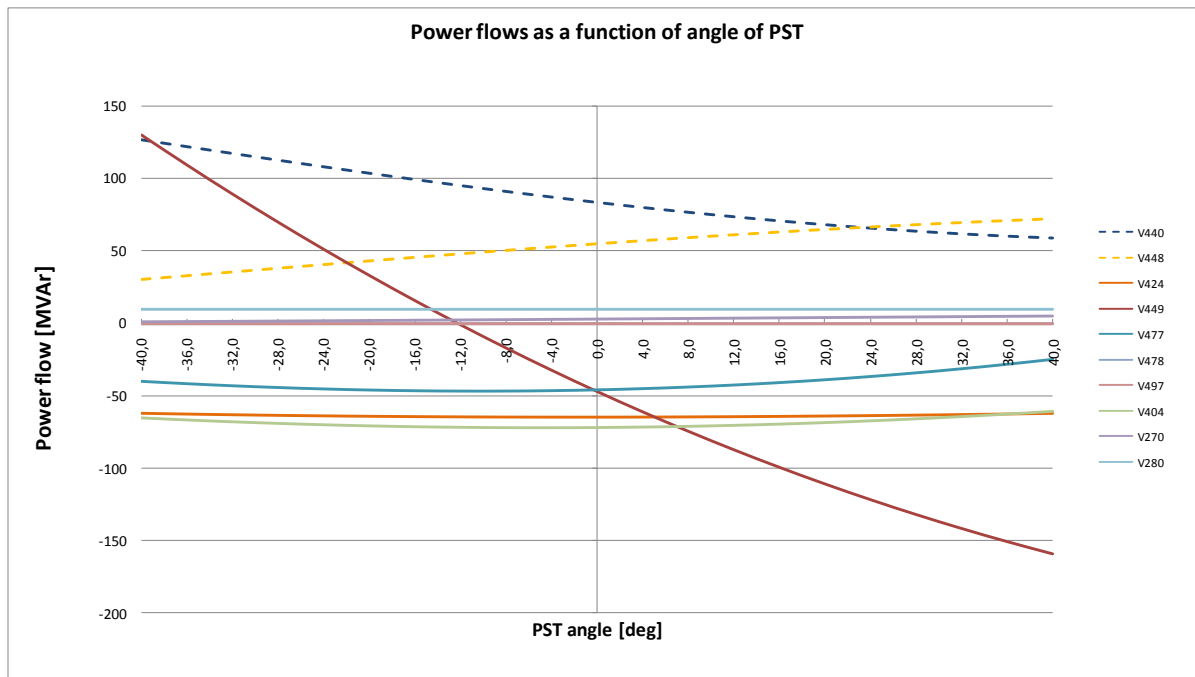


Fig. 8 Flows of re-active (non-active) power as a function of PST in line V449

## V. CONCLUSION

In this article it was shown, that PST transformer can be used to influence power flows in TS of the Slovak Republic. But it is necessary to take into account that this simulation was made without losses optimization. The main goal was to show the advantages of the use of PST.

This work was supported by Scientific Grant Agency of the Ministry of Education of Slovak Republic and the Slovak Academy of Sciences under the project VEGA No. 1/4072/07 and by the Slovak Research and Development Agency under the contract No. APVV-0385-07.

## REFERENCES

- [1] DEL VECCHIO, R. M. – POULIN, B. – FEGHALI, P. T. – SHAH, D. M. – AHUJA, R.: Transformer Design Principles With Applications to Core-Form Power Transformers, p. 499 – 504, ISBN 978-90-5699-703-8
- [2] General Electric Company: Variable Frequency Transformer™ . Dostupné na internete <[http://www.gepower.com/prod\\_serv/products/transformers\\_vft/en/downloads/vft\\_factsheet.pdf](http://www.gepower.com/prod_serv/products/transformers_vft/en/downloads/vft_factsheet.pdf)>
- [3] HINGORANI, G. N., GYUGYI, L.: Understanding FACTS. Concepts and Technology of Flexible AC Transmission Systems. New York: IEEE Press, 2000. 432 s. ISBN 0-7803-3455-8
- [4] MIHALÍKOVÁ, J.: Problém výberu simulačného nástroja pre simulačný projekt. In: Novus scientia 2007: 10. celoštátna konferencia doktorandov strojných fakúlt technických univerzít a vysokých škôl s medzinárodnou účasťou: 20.11.2007 ÚVZ Herľany, Slovenská republika. Košice : TU, 2007. s. 392-396. ISBN 978-80-8073-922-5.
- [5] ŠMIDOVIČ, R. - RUSNÁK, J.: Automatic Voltage Control in the Electric Power System of the Slovak Republic. In.: Control of Power & Heating Systems 2006: Proceedings of the 7th international conference: Zlín, Czech Republic, May 16-18, 2006. Zlín: Tomas Bata University, 2006. p. p26-1-p26-5. ISBN 80-7318-409-5.
- [6] SZKUTNIK J.: Wykorzystanie algorytmów zadań transportowych do optymalizacji dystrybucji energii elektrycznej Prace Naukowe Akademii Ekonomicznej nr. 1078, Wrocław 2005 r., s. 277-283
- [7] TŮMA, J. - MARTINEK, Z. - TESAŘOVÁ, M. - CHEMIŠINEC, I. : Security, Quality and Reliability of Electrical Energy (monografie v AJ), CONTE spol. s r.o., ČVUT Praha, ZČU v Plzni 2007, ISBN 978-80-239-9056-0 (+SW zpracování monografie pro PC)

# Video surveillance systems

<sup>1</sup>Anna KOLESÁROVÁ

<sup>1</sup>Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

<sup>1</sup>anna.kolesarova@tuke.sk

**Abstract**—This paper is concerned with video surveillance systems. The aim of the surveillance applications is to detect, track and classify targets. In this paper is described object modeling and activity analysis.

**Keywords**—Activity analysis, change detection, object modeling, video surveillance system.

## I. INTRODUCTION

Video surveillance systems are widespread and common in many environments. Video surveillance has been a key component in ensuring security at airports, banks, casinos, and correctional institutions. More recently, governments agencies, businesses, and even schools are turning toward video surveillance as a means to increase public security.

Several important research questions remain to be addressed before we can rely upon video surveillance as an effective tool for crime prevention, crime resolution, and crime protection [8].

Much of the current research in video surveillance focuses on algorithms to analyze video and other media from multiple sources to automatically detect significant events [7].

## II. VIDEO SURVEILLANCE SYSTEMS

Video surveillance is an active area of research. Object detection and tracking in video surveillance systems are commonly based on background estimation a subtraction.

Application of visual surveillance include car and pedestrian traffic monitoring, human activity surveillance for unusual activity detection, people counting, ect.

Several video surveillance products are available on the market for office and home security as well as remote surveillance [10].

The system comprises the function of object detection, tracking, recognition and classification. The problem of object detection has been tackled using statistical models of the background image [4, 10, 13], frame differences techniques or a combination of both [5].

## III. SYSTEM DESCRIPTION

The surveillance system implemented can be viewed as four independent, but interacting modules: detection, tracking, classification and recognition. The figure 1 describes the system and the interaction between its modules. To perform

the detection task, a robust real-time algorithm, suggested by T. Boulton [4] was adapted.

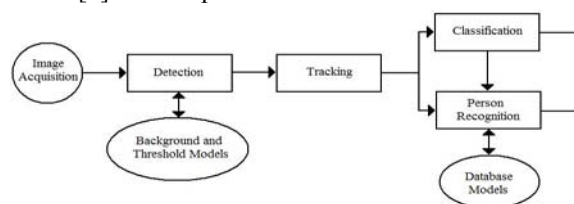


Fig. 1. System block diagram

The tracking algorithm determines the overlap between detected regions in consecutive frames, in order to link them, when no ambiguity exists.

### A. Detection

The main difficulties of such approach lie in the fact that, even in controlled environments, the background undergoes a continual change, mostly use to the existence of lighting variations and distracters (example: clouds passing by, braches of trees moving with the wind).

The system implemented uses two gray scale background models, created during a training phase. The idea is to have both a lower and a higher pixel value, contemplating this way to variations of “non target” pixels in the scene. The per-pixel threshold is then initialized to be above the difference between the two backgrounds.

### B. Tracking

The purpose of tracking is to determine the spatial-temporal information of each target present in the scene. Since the visual motion of targets is always small in comparison to their spatial extends, no position prediction is necessary to construct the strokes [13]. The association of regions and their classification is based on a binary association matrix computed by testing the overlap of regions in consecutive frames. Whenever there is a match, the stroke is updated.

### C. Classification

For the classification task three main questions must be answered, namely: which classes should be considered, which features best separate these classes and which classifiers best adapt to the previous choices? One of the main goals of the classifiers is to achieve low miss-classification probabilities while considering a wide spectrum of classes. At the same time the goal was not to consider time-dependent features,

limiting the classifier exclusively to geometric properties. In this way the resulting classifier can be used in different machines, as it is independent of the achieved frame-rate [15].

#### D. Recognition

As in the classification module, no time information is used to perform the recognition task. This recognition process is aimed at recognizing in a short term period, i.e. targets that become occluded for a few seconds or targets that merge for a few seconds and then split again. The models are characterized by the *pdf* estimates of the chosen feature space, in this color case [15].

#### IV. ACTIVITY ANALYSIS

After classifying an object, we want to determine what it is doing. Understanding human activity is one of the most difficult open problems in the area of automated video surveillance. Detecting and analyzing human motion in real time from video imagery has only recently become viable with algorithms. These algorithms represent a good first step to the problem of recognizing and analyzing humans, but they still have some drawbacks. Therefore the human subject must dominate the image frame so that the individual body components can be reliably detected [5].

#### V. OBJECT MODELING

The purpose of video surveillance systems is to monitor the activity in a specified, indoor or outdoor area. Because the image is usually captured by a stationary camera, it is easier to detect a still background than moving object. Since the cameras used in surveillance are typically stationary, a straightforward way to detect moving objects is to compare each new frame with a reference frame, representing in the best possible way the scene background [12]. The background subtraction is the higher level processing modules for object tracking, event detection and scene understanding purposes uses the results of this process [14, 16].

Background modeling is commonly carried out at pixel level [16].

#### VI. CHANGE DETECTION

For the surveillance application considered, video cameras capture images of a static scene, with illumination changes, most of the time. The entrance of an intruder into the scene can thus be detected by the changes it causes. A change detection segmentation algorithm can be used, with the changing areas typically correspondent to intruders.

The change detection algorithm implements a statistical hypothesis test to decide whether a given pixel has changed, or not and additionally, the thresholding step makes extra considerations about the differences between the changed and unchanged areas' variations, and on the size of the changed area, to achieve a better behavior for the thresholding operation [6].

#### VII. CONCLUSION

Surveillance systems significantly contribute to situation control. Such systems transform video surveillance from a data acquisition tool to information and intelligence acquisition systems. Real-time video analysis provides surveillance systems with the ability to react to an activity in real time, thus acquiring relevant information at much higher resolution [3].

Despite the importance of the subject and the intensive research done, background detection remains a challenging problem in applications with difficult circumstances, such as changing illumination, waving trees, water, rotating fans, moving shadows, inter-reflections, camouflage, occasional changes of the true background, high traffic, etc. [2].

The problem of remote surveillance has received growing attention in recent years, especially in the context of public infrastructure monitoring for transport applications, safety of quality control in industrial applications, and improved public security. The development of a surveillance system requires multidisciplinary expertise, including knowledge of signal and image processing, computer vision, communications and networking pattern recognition and sensor development and fusion [3].

#### REFERENCES

- [1] Y. Bar-Shalom, T. Formann, "Tracking and Data Association", Academic Press, 1988.
- [2] Z. Bojkovič, A. Samčovič, J. Turán, "Object Detection and Tracking in Video Surveillance systems".
- [3] Z. Bojkovič, J. Turán, "Key Challenges in Video Based Surveillance Systems", Košice, Sept. 2005.
- [4] T. Boulton, et al., "Into the Woods: Visual Surveillance of Noncooperative and Camouflaged Targets in Complex Outdoor Settings", in Proceeding of the IEEE, vol. 89, no. 10, Oct. 2001.
- [5] R. Collins, et al., "A System for Video Surveillance and Monitoring", CMU-RI-TR-00-12, 2000.
- [6] P. L. Correia, F. Pereira, "Change Detection – Based Video Segmentation for Surveillance Applications", VIAMIS, 2004.
- [7] R. Cucchiara, "Multimedia Surveillance systems", Proc. ACM Workshop on Video Surveillance and Sensor Networks, pp. 3 -10, 2005.
- [8] Z. Dimitrijevic, G. Wu, E. Chang, "A Multi-Sensor Fusion and Mining System", Proceeding of the 2nd Usenix FAST, March 2003.
- [9] A. Elgammal, D. Harwood, L. A. Davis, "Non-Parametric Model for Background Subtraction", ICCV'99, 1999.
- [10] H. Haritaoglu, "Hartwood and Devis, W4: Real Time Surveillance of People and their Activities", IEEE Trans. Pattern Anal. Machine Intell., vol. 22, no. 8, Aug. 2000, pp. 809 - 830.
- [11] T. Horprasert, D. Harwood, L. A. Davis, "A Statistical Approach for Real-Time Robust Background Subtraction and Shadow Detection", ICCV'99 Frame Rate Workshop, 1999.
- [12] C. Ianasi, V. Gui, C. I. Toma, D. Pescaru, "A Fast Algorithm for Background Tracking in Video Surveillance, Using Nonparametric Kernel Density Estimation", Elec. Energ. vol 18, No 1, April 2005, 127-144 .
- [13] S. McKenna, et al., "Tracking Groups of People", CVIU 80, pp. 42 - 56, 2000.
- [14] D. Murray, A. Basu: "Motion Tracking with an Active Camera", IEEE Trans. Pattern Recogn. And Machine Intell, vol. 19, no. 5, pp. 449 - 454, May 1994.
- [15] R. J. Oliveira, P. C. Riberio, J. S. Marques, J. M. Lemos, "A Video System for Urban Surveillance: Function Integration and Evaluation", VIAMIS, 2004.
- [16] K. R. Rao, Z. S. Bojkovič, D. A. Milovanović, "Introduction to Multimedia Communications: Applications, Middleware, Networking", Wiley, 2005.
- [17] K. R. Rao, Z. S. Bojkovič, D. A. Milovanović, "Multimedia Communication Systems: Techniques, Standards and Networks", Prentice-Hall PTR, New Jersey, 2002.



# Fuzzy Vector Control of AM

Michal KOVÁČ

<sup>1</sup>Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

michal.kovac.83@gmail.com

**Abstract** — Considered in the present paper is applying artificial intelligence to vector-oriented controlling of a 0.75kW squirrel-cage asynchronous motor. The very control is realized through a pair of Takagi-Sugeno fuzzy controllers. Described below will be steps of designing individual fuzzy controllers, and also a comparison of their simulation results with those of a conventional state controller. Subsequently, presented will be the principle governing adaptation of the fuzzy controller to changes of the drive parameters' changes caused by heating-up and by mechanical alterations to the drive.

**Keywords** — Asynchronous motor (AM), Vector control (VR), Fuzzy controller (FR).

## I. INTRODUCTION

Nowadays, the asynchronous motor presents one of the most widely used drives in practice. It features a number of advantages such as simple structure, low price, reliability, simple and practically least maintenance. In the past it was most widely utilized in the domain of uncontrolled drives whereas quality of the asynchronous motor state control was not attaining qualities of the then dominant controlled DC independently excited drives. The vector control theory that was publicized in the 70's of the past century for the first time, which reversed the situation. Introduced were modern electrical controlled drives, and recently these are most commonly realized. A new line-up of modern electrical controlled drives appeared on the scene, presently most commonly being realized by the squirrel cage asynchronous motor. This area of electrical drives is characteristic by small time constants and by intricate or possibly unknown mathematical description due to profound non-linearity in the electromagnetic part of the drive or due to the technology and/or non-measurable breakdowns of the electrical drive. Since the birth of the theory developed and published were a multitude of traditional controllers for the type of control. Nonetheless, their most outstanding disadvantage is seen in the must to know any parameters and the drive math model, which is consequently simplified and linearized. For the reason, applied in the recent years came on the field the fuzzy set theory as the means of modelling uncertainty in natural language, which is the basis of human thinking. The approach is more approximate than exact and has a number of undeniable advantages that support the assumption that

artificial intelligence (UI) methods can be applied in the form of fuzzy logic controllers (FLC) even in the drive practice.

## II. AM VECTOR CONTROLLING

Vector AM controlling has highly favourable properties as in steady so in transient states. It is based on the analogy with DC separately excited motor the momentum of which is achieved by magnetic excitation flux and by current. In the vector control, the stator current is distributed between two vertical constituents in the x, y coordinate system. Used in this case is direct vector-oriented control of the rotor flux, structural block diagram of which is shown in the figure below.

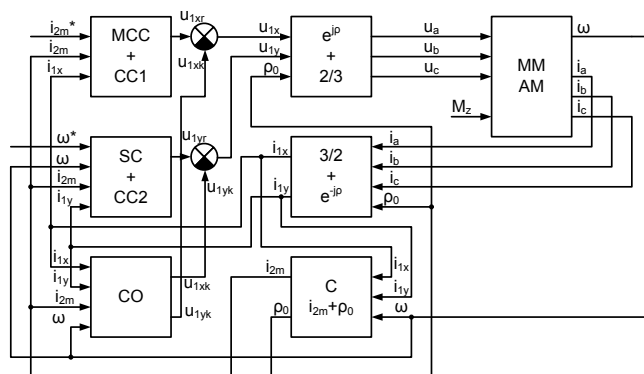


Fig.1 Structured block diagram of the AM vector control  
MCC- magnetizing current controller, CC1-current controller 1, SC-speed controller, CC2-current controller 2, MM-mathematical model, C-calculation, CO-compensation

Output values of individual controllers are voltages expressed in the x, y coordinate system that rotates by  $\omega_k$  asynchronous speed. These quantities are transformed into a system of firm stator coordinates  $\alpha, \beta$ , and are subsequently transformed from two-phase coordinate system  $\alpha, \beta$  into three-phase coordinate system a,b,c, and which still subsequently enter the AM mathematical model. Outputs of this are currents  $i_a, i_b, i_c$  that are back-transformed into  $i_x, i_y$ . Calculated from these currents and from the async motor speed are the actual magnetizing current  $i_{2m}$  and angle  $\rho_0$  that is necessary for individual transformations.

Vector-oriented control of the rotor flux was realized using fuzzy controllers, waveforms of which were compared against traditional conventional state control designed using the pole determination method for two input and one output quantities.

### III. FUZZY CONTROLLER DESIGN

Used for controlling were two Takagi-Sugeno fuzzy controllers the structure of which, same as the operation principle are very similar to those of Mamdaniho fuzzy controller. Moreover, the controller lowers high hardware requirements mostly due to de-fuzzification.

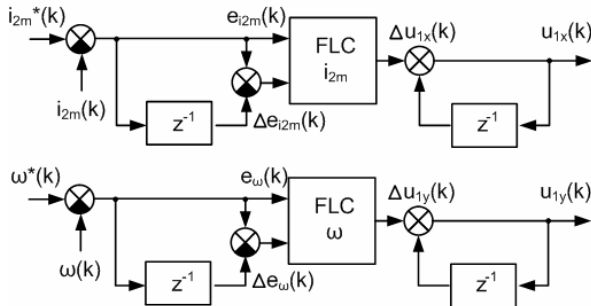


Fig.2 Diagram of the discrete PI fuzzy controller FLC-fuzzy logic controller

One Takagi-Sugeno fuzzy controller controlled the magnetizing flux, and the other one controlled speed. Controllers had two inputs and one output. Used for the first controller input was control error  $e$ , which informs one on the difference between the desired (\*) and actual value obtained from the system. At controlling an asynchronous motor is it necessary to know also its dynamics and direction. The necessity was met by the second controller input – derivation of deviation  $\Delta e$ . The fuzzy controller output is value of action quantity  $\Delta u$ .

PI fuzzy controller with  $T=0.001s$  sampling period decides based on the information on the control error magnitude and on its derivation and determines whether the action quantity will be lowered, increased or let on its original value.

$$\Delta u_{(k)} = K_P \cdot \Delta e_{(k)} + K_I \cdot e_{(k)} \quad (1)$$

Used with both controller types were triangular membership functions whereas they can be analytically described with utmost ease.

The magnetizing current Fuzzy controller was preset for five states of the control error and another five states for its derivation (Fig. 3). Individual states were marked as (**BN** – big negative, **MN** – mean negative, **Z** – zero, **MP** – mean positive & **BP** – big positive).

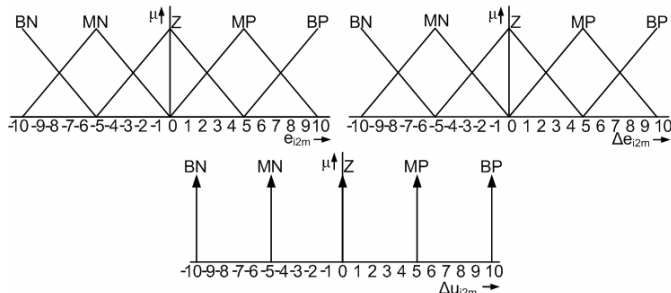


Fig.3 Membership functions for the magnetizing current Takagi-Sugeno fuzzy controller

In the case of the SPEED fuzzy controller it proved to be necessary to increase the number of possible states to enhance resolving ability in close vicinity of steady state condition (Fig. 4). The control error had seven differing states, and its derivation had five (**BN** – big negative, **MN** – mean negative, **SN** – small negative, **Z** – zero, **SP** – small positive, **MP** – mean positive a **BP** – big positive).

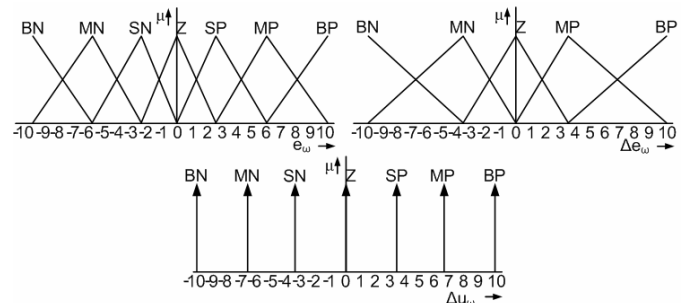


Fig.4 Membership function of the Takagi-Sugeno fuzzy speed controller

Applied was the **min-max** inference according to individual rules. In case of the magnetizing current controller it meant 25 rules and in case of the speed controller it meant 35 rules (Fig.5) of the type:

**PI controller:**

**IF**  $e_{(k)}$  is ..... **AND**  $\Delta e_{(k)}$  is ..... **THEN**  $\Delta u_{(k)}$  is .....

a,						b,							
$\Delta e \backslash e$	LN	MN	Z	MP	LP	$\Delta e \backslash e$	LN	MN	SN	Z	SP	MP	LP
LN	LN	LN	LN	Z	MP	LN	LN	LN	LN	MN	Z	SP	MP
MN	LN	LN	MN	MP	LP	MN	LN	LN	MN	SN	SP	MP	LP
Z	LN	MN	Z	MP	LP	Z	LN	MN	MN	Z	MP	MP	LP
MP	LN	MN	MP	LP	LP	MP	LN	MN	SN	SP	MP	LP	LP
LP	MN	Z	LP	LP	LP	LP	MN	SN	Z	MP	LP	LP	LP

Fig.5 Determination table for Takagi-Sugeno fuzzy controller of magnetizing current (a,) and of speed (b,)

Attained at the Takagi-Sugeno fuzzy controller is the overall sharp number  $\Delta u$  as weighted average  $U_i$  with forces of individual rules  $\mu_i$  at the total number of rules =  $n$ .

$$\Delta u = \frac{\sum_{i=1}^n \mu_i \cdot U_i}{\sum_{i=1}^n \mu_i} \quad (2)$$

### IV. SIMULATION RESULTS

Compared in the paragraph below will be waveforms of individual drive quantities derived from using both Takagi-Sugeno fuzzy controller and a traditional state controller, respectively, with preset damping  $d=0.85$  and control time  $t_r=0,1s$ . Simulations were effected within Simulink (Matlab) application using an async motor model with 0.75kW output, which was in the period from 0.4s to 1.1s loaded with torque  $M_z=5Nm$ .

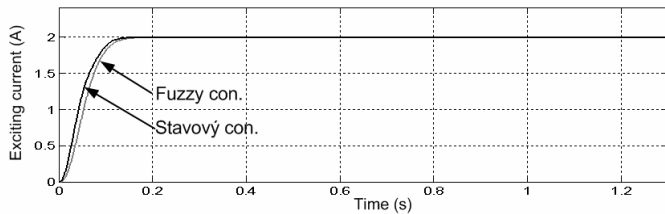


Fig.6 Waveforms of magnetizing current  $i_{2m}$  of fuzzy & state cont

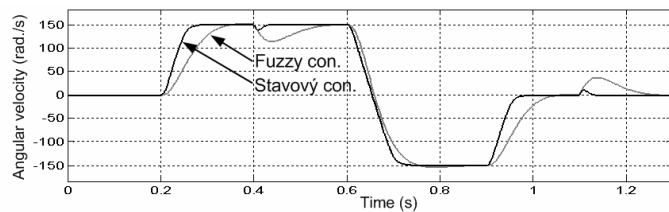


Fig.7 Waveforms of angular speed  $\omega$  of fuzzy & state cont

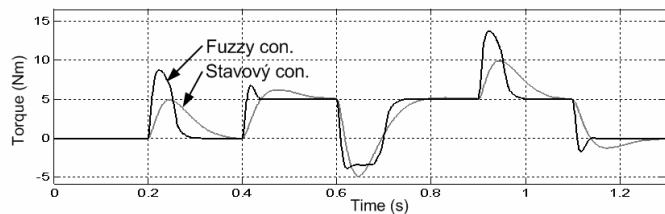


Fig.8 Waveform of motor speed  $M_m$  of fuzzy & state cont

As can be seen on the waveforms above the fuzzy controller responded more promptly to transition states, and hence sooner regulated the control error, whilst the torque did not exceed 2.5-multiple of the rated torque ( $M_N=5,19Nm$ ).

Shown in Figs. 8 & 9 are waveforms of control error of individual controller, its derivation and subsequent control intervention of the Takagi-Sugenov fuzzy controller.

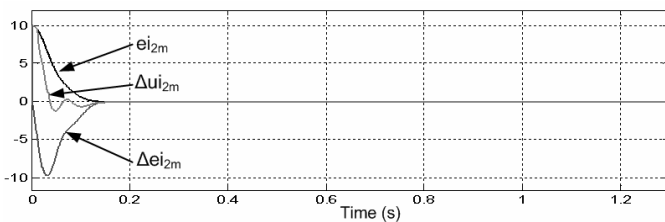


Fig.9 Waveforms of quantities entering and leaving the magnetizing current controlling fuzzy controller

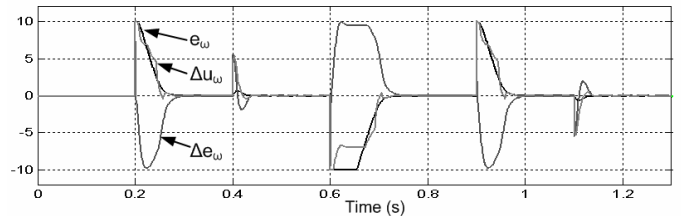


Fig.10 Waveforms of quantities entering and leaving RPMS controlling fuzzy controller

The above waveforms of controlled quantities were always of a-periodic nature. A proof of it can be seen on dependence of the speed controlling controller's control error upon its derivation (Fig. 10). At the drive start-up to the rated angular speed the vector is in the fourth quadrant, and at its subsequent reversal in the second quadrant. Changeover between the quadrant is caused by a change in the disturbance variable which is in our case represented by load torque  $M_z$ .

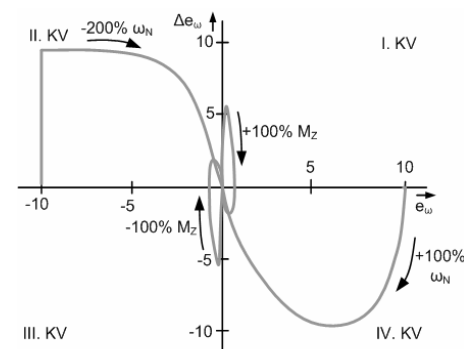


Fig.11 Waveform of the control error  $e$  and its derivation  $\Delta e$  in the fuzzy speed controller

During operation, the electric current flow causes the drive to heat up. This results in the fact that the async motor's parameters change in time. At the stator winding resistance this may be up to 20% as compared with rated values and at the rotor resistance it may be even up to 50%. Shown in the following figures (11 & 12) are waveforms of the angular speed under such influences.

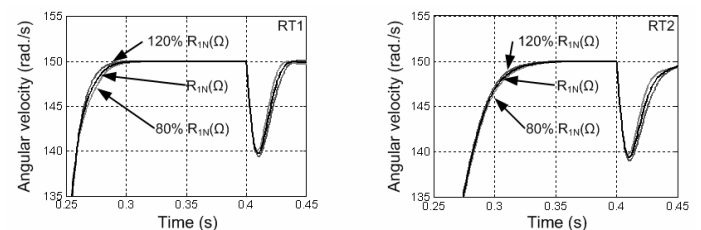


Fig.12 Waveforms of the angular speed at change of the stator winding resistance  $R_1$

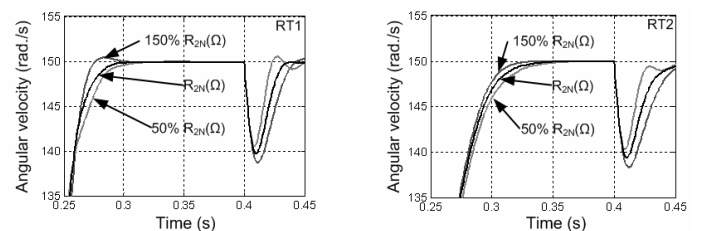


Fig.13 Waveforms of the angular speed at change of the rotor winding resistance  $R_2$

Shown in the left column are the angular speed waveforms when using the speed controlling fuzzy controller, which was deciding based on the determination table RT1 rules of which are specified in Fig. 4. Shown in the right column are waveforms of the angular speed in case of which rules of the determination table RT2 were altered so that the drive would reflect a-period waveforms of the quantity controlled even under the most adverse conditions. The method of appropriate selection of determination table with adequately set rules allows us to set up a drive with adaptive control structure.

The angular speed transition states are not influenced by changes in the drive electrical parameters only, but also by mechanical changes. In concern is especially the moment of inertia that can fluctuate during operation. It does not matter whether in concern are winders or un-winders. Even in this case is adaptation of the speed controller based on change of rules set in the determination table the simplest way. Consequently, ensured can be comparable angular speed waveform at the rated moment of inertia  $J_{ZN}$  also at its subsequent increase by 200% (Fig. 13).

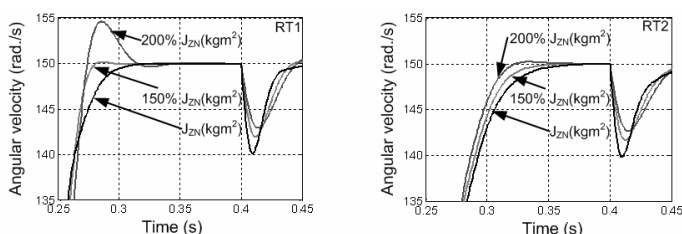


Fig.14 Waveforms of the angular speed at changes in the moment of inertia  $J_z$

## V. CONCLUSIONS

Contrary to the field of the traditional controller, presently there is no precise methodology for techniques involved in exact designing of individual parameters of the fuzzy controller. Seen from the fuzzy logic theory perspective, this could not be even possible whereas the very title “fuzzy” stands for something approximate, unclear, imprecise. Majority of the fuzzy controller designs still rests upon experience of the expert who is able to describe functioning of the system in words. In concern is the so-called “subjective assessment and remuneration method” that makes him or her face a number of challenges concerning the appropriate sampling period based on the number of fuzzy sets that describe a scope of individual input and output values, on the number and shape of the membership function to achieve as simple controller as possible and that, last but not least, ensure that preserved will be high quality of controlling, dynamics and robustness.

Proper answers to the above-posted issues are based on the experience of the expert designing and fine-tuning the fuzzy controller most often applying the trial and error method. Correct tuning of the controller is time consuming, and its quality does not have to meet the criteria imposed on quality of control requested for any states the drive can be in.

On this stage adaptive fuzzy controllers grab the steering wheel. The expert proposes initial distribution of the membership functions and the rules. Subsequent manual fine-tuning is achieved by an artificial intelligence method regardless of the fact that in concern are artificial neural

networks or genetic algorithms. The latter ones, based on the experience of the expert on behaviour of a system at changes to its parameters, will learn what fuzzy controller adaptation is the most advantageous one at the moment. The advantage of this approach is seen in the fact that it is not longer needed to be aware of all system parameters and to design a new controller whenever the drive parameters change. Fuzzy controller adapts to changing conditions all by itself, which is impossible to expect from controllers designed by conventional methods.

Such an online learning may be divided into two parts; the first one is fine-tuning of the drive prior to its introducing into operation, and the second one is fine-tuning of the very drive once it is put into operation where the drive adapts to those parameters of the drive system that were neglected or linearised at designing.

## ACKNOWLEDGMENT

The assignment has been solved within the grant VEGA No.1/4076/07.

## REFERENCES

- [1] TIMKO, J. – ŽILKOVÁ, J. – GIROVSKÝ, P.: Electrical drives. TU Košice, ISBN 80-8073-529-8, pp 149, Košice 2007.
- [2] BRANDŠTETTER, P. – ŠTEPANEC, L.: Fuzzy Logic Control of Induction Motor Drive. IWCIT'01, VŠB-TU Ostrava, 2001, ISBN 80-7078-907-7.
- [3] CIBUĽA, Ľ. - TIMKO, J. - ŽILKOVÁ, J. - GIROVSKÝ, P. : Direct torque control of the asynchronous motor. In: Journal of Computer Science and Control Systems. (2008), p.18-21. ISSN 1844-6043.
- [4] TIMKO, J. – ŽILKOVÁ, J. - ,GIROVSKÝ, P. : Shaft sensorless vector control of an induction motor. In: Acta Technica ČSAV. Vol.52, no.1 (2007), p.81-91. ISSN 0001-7043.
- [5] VITTEK, J. et al. : Comparison of sliding mode and forced dynamics control of electric drive with flexible coupling employing PSMM. In : IEEE International Conference on Industrial Technology IEEE ICIT 2008:21-24 April 2008. Piscataway: IEEE 2008, p.1-6. ISBN 978-1-4244-1706-3.
- [6] ŽILKOVÁ, J. – TIMKO, J. – GIROVSKÝ, P. : Nonlinear system control using neural networks. In: Acta Polytechnica Hungarica. Vol.3, no.4 (2006), p.85-94. ISSN 1785-8860.

# Discription, Effects, Monitoring and Trends of Partial Discharges

<sup>1</sup>Milan KVAKOVSKÝ, <sup>2</sup>Vieroslava ČAČKOVÁ, <sup>3</sup>Lýdia DEDINSKÁ

<sup>1</sup> Department of Electric Power Engineering, FEI TU of Košice, Slovak Republic

<sup>2</sup> Department of Electric Power Engineering, FEI TU of Košice, Slovak Republic

<sup>3</sup> Department of Electric Power Engineering, FEI TU of Košice, Slovak Republic

<sup>1</sup>milan.kvakovsky@tuke.sk, <sup>2</sup>vieroslava.cackova@tuke.sk, <sup>3</sup>lydia.dedinska@tuke.sk

**Abstract**— Partial discharge (PD) is a localized dielectric breakdown of a small portion of a solid or liquid electrical insulation system under high voltage stress. While a corona discharge is usually revealed by a relatively steady glow or brush discharge in air, partial discharges within an insulation system may or may not exhibit visible discharges, and discharge events tend to be more sporadic in nature than corona discharges. PD usually begins within voids, cracks, or inclusions within a solid dielectric, at conductor-dielectric interfaces within solid or liquid dielectrics, or in bubbles within liquid dielectrics. Since discharges are limited to only a portion of the insulation, the discharges only partially bridge the distance between electrodes. PD can also occur along the boundary between different insulating materials.

**Keywords**—Partial discharge, on-line measurement, PD types, paper – oil insulation.

## I. INTRODUCTION

PD usually begins within voids, cracks, or inclusions within a solid dielectric, at conductor-dielectric interfaces within solid or liquid dielectrics, or in bubbles within liquid dielectrics. Since discharges are limited to only a portion of the insulation, the discharges only partially bridge the distance between electrodes. PD can also occur along the boundary between different insulating materials.

Partial discharges within an insulating material are usually initiated within gas-filled voids within the dielectric. Because the dielectric constant of the void is considerably less than the surrounding dielectric, the electric field (and the voltage stress) appearing across the void is significantly higher than across an equivalent distance of dielectric. If the voltage stress across the void is increased above the corona inception voltage (CIV) for the gas within the void, then PD activity will start within the void.

Once begun, PD causes progressive deterioration of insulating materials, ultimately leading to electrical breakdown. PD can be prevented through careful design and material selection. In critical high voltage equipment, the integrity of the insulation is confirmed using PD detection equipment during the manufacturing stage as well as periodically through the equipment's useful life. PD prevention and detection are essential to ensure reliable, long-

term operation of high voltage equipment used by electric power utilities [3].

## II. PD TYPES AND THEIR GENERATION SYSTEMS HV

For the purpose of this study, the following classification of PD is assumed, which enables to select the most important PD type that frequently occur in insulation of electric power facilities:

1. point-plane type discharges in oil,
2. surface discharges in oil,
3. gas bubble discharges in oil,
4. discharges in particles moving in oil.

Model systems have been constructed for each of the specified PD to be generated. Simplified schematics of these systems are respectively shown in Figures 1 through 3. The constructed spark gaps enable generation of PD of apparent charge ranging from ~ 3 to ~ 8000 pC. This range covers both those PD which are recognized as harmless to very harmful in insulation systems. The spark gaps are submerged in transformer oil and supplied from a standard HV test system [4].

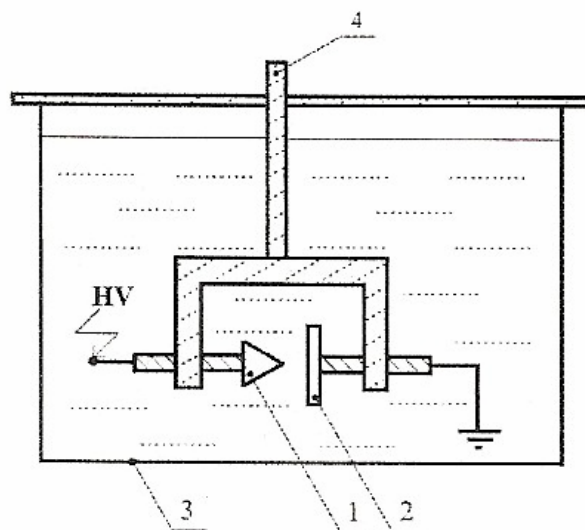


Fig. 1. Schematic of a 'point to ground plane' spark gap 1: point electrode, 2: flat electrode, 3: insulation oil tank, 4: structure supporting the gap [4].

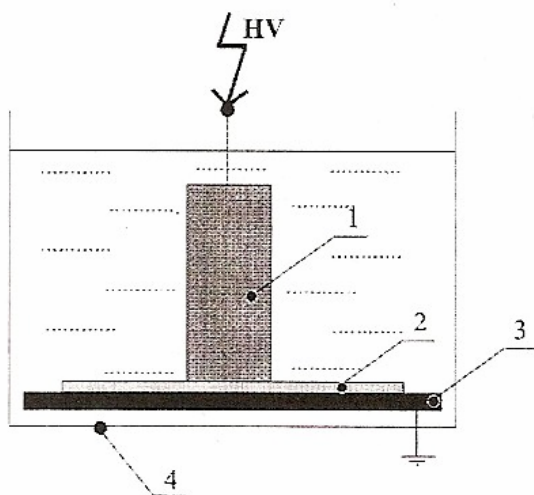


Fig. 2. Schematic of a 'cylinder to ground plane' spark gap, with a pressboard insulation, to generate surface PD. 1: cylinder electrode, 2: pressboard insulation, 3: flat grounded electrode, 4: insulation oil tank [4].

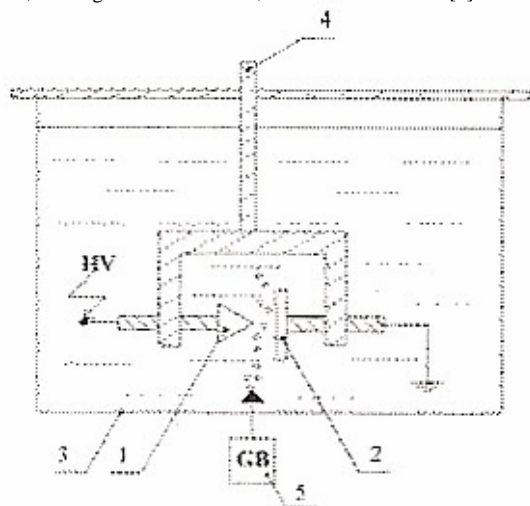


Fig. 3. Schematic of a 'point to ground plane' spark gap to generate gas bubble discharges in oil 1: point electrode, 2: flat electrode, 3: insulation oil tank, 4: holder for the gap, 5: generator of air bubbles [4].

### III. EFFECTS OF PARTIAL DISCHARGE IN INSULATION SYSTEMS

PDs are localized ionization within electrical insulation that is caused by a high electrical field. They occur in part of the insulation system and are limited in extent, so they do not immediately cause full insulation breakdown.

PD can occur in a gaseous, liquid or solid insulating medium. It is often initiated within gas voids enclosed in solid insulation, or in bubbles within a liquid insulating material, such as voids in an epoxy insulator, or gas bubbles dissolved within transformer oil.

As the gas within the void has a dielectric constant much less than the surrounding material, it experiences a significantly higher electric field. When this becomes high enough to cause electrical breakdown in the gas, a partial discharge occurs. PD can also occur along the surface of solid insulating materials if the surface tangential electric field is high enough to cause a breakdown along the insulator surface. This phenomenon commonly manifests itself on overhead line

insulators, particularly on contaminated insulators during days of high humidity. Overhead line insulators use air as their insulation medium.

The effects of PD within high voltage cables and equipment can be very serious, ultimately leading to complete failure. The cumulative effect of partial discharges within solid dielectrics is the formation of numerous, branching partially conducting discharge channels, a process called treeing. Repetitive discharge events cause irreversible mechanical and chemical deterioration of the insulating material. Damage is caused by the energy dissipated by high energy electrons or ions, ultraviolet light from the discharges, ozone attacking the void walls, and cracking as the chemical breakdown processes liberate gases at high pressure. The chemical transformation of the dielectric also tends to increase the electrical conductivity of the dielectric material surrounding the voids. This increases the electrical stress in the (thus far) unaffected gap region, accelerating the breakdown process. A number of inorganic dielectrics, including glass, porcelain, and mica, are significantly more resistant to PD damage than organic and polymer dielectrics.

In paper-insulated high-voltage cables, partial discharges begin as small pinholes penetrating the paper windings that are adjacent to the electrical conductor or outer sheath. As PD activity progresses, the repetitive discharges eventually cause permanent chemical changes within the affected paper layers and impregnating dielectric fluid. Over time, partially conducting carbonized trees are formed. This places greater stress on the remaining insulation, leading to further growth of the damaged region, resistive heating along the tree, and further charring (sometimes called *tracking*). This eventually culminates in the complete dielectric failure of the cable and, typically, an electrical explosion [2].

PD dissipate energy, generally in the form of heat, but sometimes in as sound and light as well, like the hissing and dim glowing from the overhead line insulators. Heat energy dissipation may cause thermal degradation of the insulation, although the level is generally low. For high voltage equipment, the integrity of the insulation can be confirmed by monitoring the PD activities that occur through the equipment's life. To ensure supply reliability and long-term operational sustainability, PD in high-voltage electrical equipment should be monitored closely with early warning signals for inspection and maintenance.

### IV. SELECTION OF BASIC PD FORMS THAT CAN OCCUR IN PAPER-OIL INSULATION

Different ways of PD classification can be found in dealing with electrotechnical materials and high-voltage technology. It is possible, based on energy criterion, to divide discharges into self-maintained and not self-maintained ones. They can be classified according to PD intensity change in time. In this way PDs can be divided into stable, unstable and decaying discharges due to electric field activity. They can be also divided according to geometry of damage and dielectric type in which they are generated. From this point, of view discharges are divided into PDs generated in gases, solid and liquid dielectrics.

During the research investigations carried out, authors adopted the division which makes it possible to isolate the most significant PD forms that can occur in paper-oil insulation systems of electric power appliances. In order to systematize the terms used in this paper, the notion of a 'class' was introduced, which referred to a definite basic PD form. Adopting eight PD forms for analysis, the following classes have been defined: *class 1 - discharges in the point-point system in oil*, which can be related to PDs generated by insulation damage of two neighboring turns of winding of a transformer; *class 2 - discharges in the point-point system in oil with gas bubbles*, which can reflect PDs in gassy oil and which are caused by insulation damage of two neighboring turns of winding of a transformer; *class 3 - discharges in the point-plane system in oil*, which can model PDs occurring between a damaged part of transformer winding insulation and earthed flat parts (tub elements); *class 4 - discharges in the surface system of two flat electrodes with paper-oil insulation between them*, the most common PD form occurring in the so-called triple point, in which an electrode surface is in contact with solid and liquid dielectrics; *class 5 - discharges in the surface system of one flat electrode and one multipoint electrode*, with paper-oil insulation between them, a different electric field intensity distribution in comparison with the surface system with two flat electrodes; *class 6 - discharges in the multipoint-plane system in oil*, which can model PDs occurring between a multipoint damage of transformer winding insulation and earthed flat parts (tub elements); *class 7 - discharges in the multipoint-plane system in oil with gas bubbles*, which can be connected with PDs occurring between a multipoint damage of transformer winding insulation and earthed flat parts (tub elements), but in oil with gas molecules; *class 8 — discharges on indeterminate potential particles moving in oil*, which can model PDs occurring in oil containing particles of cellulose fibers<sup>15</sup> formed in the process of a gradual degradation of paper-oil insulation caused by aging processes.

For each PD form selected, spark-gaps which modeled them were constructed and placed in a transformer tub filled with insulation oil. The spark-gaps constructed make generation of PDs of apparent charge  $Q_p$  in the range from a few pC to about 8000 pC possible. Such a value range of the apparent charge is significant from technical point of view, because this range contains PDs which are considered to be harmless, of indefinite harmfulness and of a confirmed harmful impact on insulation systems [5].

## V. PD MONITORING AND TREND ANALYSIS

### A. Introduction

PD testing is particularly important where HV plant has a high criticality.

This may be due to its age, historical failures or the consequences of its failure (position in the network).

Identification of the 'critical plant' within the plant owner's HV network can be achieved quickly and easily using IPEC's on-line PD Testing technology to provide an 'early warning system' for incipient faults. On-line PD monitoring allows for

analysis trends in PD activity to be observed over time. This may reveal correlation with environmental (temperature, humidity etc) or service conditions (changes in load etc). As PD activity is often present well in advance of insulation failure it is possible by observing its development that strategic decisions can be made about refurbishing and renewal programs [1].

### B. The benefits of on-line partial discharge field measurements

- It is truly a predictive test, indicating insulation degradation in advance of the failure.
- It is a nonintrusive test, requiring no interruption of service and is performed under normal operating voltage and load.
- It is a nondestructive test; it does not test to failure or adversely affect the equipment under test.
- It need not use any overvoltages, thereby not exposing the tested equipment to higher voltage stresses than those encountered under normal operating conditions.
- Trending can be accomplished by storing results to allow comparison with future tests.
- In many instances the site of the partial discharge occurrence can be located within the test object, so the localised problem can be repaired.
- The cost to perform a PD survey is relatively inexpensive compared with off-line testing, allowing annual surveys to be performed economically at most facilities.

### C. Examples of the HV plant that can be tested

- Cables and Cable Accessories (terminations and joints)
- Switchgear (AIS and GIS)
- Instrument transformers (voltage and current)
- Power transformers and bushings
- Motors and generators
- Surge arrestors
- Capacitors

## VI. CONCLUSIONS

High requirements as to the supply quality, imposed by a competitive electric energy market, make it necessary for distributing companies to maintain the highest reliability possible of appliances that are parts of the electric power system. Minimizing often high financial loss incurred due to a failure of electric power objects and also high costs related to electric power not delivered to consumers can be achieved through an effective power appliances. A number of discharge detection schemes have been invented since the importance of PD was realized early in the last century. Partial discharge currents tend to be of short duration and have rise times in the

nanosecond regime. On an oscilloscope, the discharges look like randomly occurring 'spikes' or pulses. The usual way of quantifying partial discharge magnitude is in picocoulombs.

#### REFERENCES

- [1] High Voltage Engineering Fundamentals, E.Kuffel, W.S. Zaengl, pub. Pergamon Press. First edition, 1992 ISBN 0-08-024213-8
- [2] Engineering Dielectrics, Volume IIA, Electrical Properties of Solid Insulating Materials: Molecular Structure and Electrical Behavior, R. Bartnikas, R. M Eichhorn, ASTM Special Technical Publication 783, ASTM, 1999
- [3] Engineering Dielectrics, Volume I, Corona Measurement and Interpretation, R. Bartnikas, E. J. McMahon, ASTM Special Technical Publication 669, ASTM, 1979, ISBN 0-8031-0332-8B. Smith, "An approach to graphs of linear forms (Unpublished work style)," unpublished.
- [4] T. Boczar, "Identification of a specific type of partial discharges form acoustic emission frequency spectra ", IEEE Trans. Dielectr. Electr. Insul., Vol. 11, pp. 598 – 606, 2001
- [5] T. Boczar and D. Zmarzly, "Application of wavelet analysis to acoustic emission pulses generated by partial discharges ", IEEE Trans. Dielectr. Insul., Vol. 8, pp. 433 – 449, 2004



# Basic aspects of smart metering system

<sup>1</sup>František Lizák

Department of electric power engineering, FEI TU of Košice, Slovak Republic.

<sup>1</sup>frantisek.lizak@tuke.sk

**Abstract** - The purpose of this paper is to provide customers and participants of electricity market basic information about new effective tools for decreasing technical and non-technical losses and also encourage low voltage customers to use electricity more efficiently and reduce their costs of energy, reduce their electricity bill.

**Keywords** – smart meter, low voltage, remote meter reading, energy efficiency, theft, tariffs, emissions

## I. INTRODUCTION

This is a general definition for an electronic device that can measure the consumption of energy (electricity or gas) adding more information than a conventional meter (price schemes, interval data, quality of supply, etc...), and that can transmit data using a form of electronic communication. Similar meters, usually referred to as ‘time-of-use’ or ‘interval’ meters, have existed for years, but smart meters usually involve a different technology mix such as automated meter reading, automated meter management and a different application mix such as domotics, value-added services, etc...

## II. TECHNICAL BACKGROUND

Smart meters are a modern system of metering that has the potential to change significantly the way in which many of the activities of electricity supply are undertaken. The metering can include a range of functions and capabilities including [1]:

- Display and recording of real time information
- An internal memory that will store previous patterns of consumption and allow trends and comparisons to be discerned
- Distinguish between the import and export of electricity
- Undertake two-way communication with a central processor that would permit the use of the data for billing and other purposes, such as risk management
- Receive instructions to switch designated circuits, or all supplies to premises.
- Power quality measurement (incl. continuity of supply and voltage quality).

Smart metering needs to be viewed as a system rather than a single device. Many of the benefits that are attributed to smart meters will require not only measurement and collection of data but also the ability to access, process, store, and retrieve that data. Thus in considering the adoption of smart meters

the limitations of other systems in the value chain need to be taken into account in determining the functionality that can be ultimately utilized [1].

## III. METER DESIGN AND DATA RETRIEVAL

**Meter design:**

Currently there is no universally adopted specification for the design of a smart meter. After carefully assessing the relevant costs and benefits, it is important to define certain minimum smart meter functionality. In order to allow for economic optimal solution and technical innovation, the individual meter service provider should be left to decide on the technical solution to fulfil the required functionality. The following main functionalities should be carefully considered:

- Remote meter reading
- Load profile data
- On demand meter data access for customer
- On demand meter data access for 3rd party
- Provision of variable time-of-use tariffs (time bands)
- Remote meter management
- Remote demand reduction and connection/disconnection
- Main communication port (e.g. GPRS, GSM, PLC, etc).
- Price signal to customer

The above list is not exhaustive but reflects the current state-of-the-art technology.

**Architecture:**

The general configuration of the communication arrangements for gathering data recorded by a smart meter is for there to be local area network (LAN) over which data can be fed to a ‘concentrator’ and then for a wireless network (WAN) to communicate data back to a central processor. Having gathered the data a number of possibilities then exist for how the processed information could be used by the supplier and returned to the customer. The most economic arrangement and technology for this general design will need to be identified and its application linked to particular geographic areas [2].

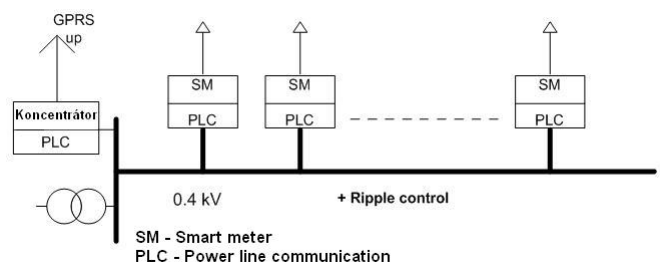


Figure 1: General conception of communication

#### IV. NETWORK FUNCTIONALITY

- **Remote meter reading.**

An obvious attribute of a smart metering system is its ability to dispense with conventional meter reading practices. The recording and remote collection of interval consumption data presents an opportunity for its use for a variety of applications including the accommodation of import and export for micro generation installations.

- **Remote connection/disconnection.**

The ability to allow electricity to be cut off or restored remotely would obviate need for visits by field staff. It may also be appropriate to contemplate functionality whereby supplies could be given at a reduced capacity which might support circumstances where there were debt collection problems.

- **Fault monitoring.**

Networks might also consider the prospects for detecting faults by monitoring the status of smart meters.

- **Theft.**

Smart meters have the ability to be equipped with tamper detection facilities. Software that can detect abnormal trends in consumption would also be a functionality that would help prevent theft and measure losses more accurately.

- **Quality of supply measurement.**

Smart meters could also be used to record outages and supply quality such as voltage abnormalities at different locations.

#### V. SUPPLIER FUNCTIONALITY

- **Time of Use tariffs.**

Assessing the impact of time of use tariffs on customer behaviour will be a key requirement of this phase of the implementation plan. The ability of tariff design to impact customer behaviour is central to encouraging energy efficiency and reduced emissions. However, the effectiveness of tariff design in achieving this goal needs to be tested.

- **Meter displays.**

Although often assumed, smart meters do not naturally provide a display of information to the customer. The nature of such displays and the information customers would find useful also requires further research. Local processing capability, such as the inclusion of tariff rates for individual suppliers, or the ability to calculate trends and predictions is a further area where there are a range of possibilities.

- **Pre-payment.**

The replacement of credit with pre-payment arrangements could have a significant impact on the bad debts seen by suppliers and thus on the costs borne by all customers. Smart meters may offer opportunities for more frequent billing of customers and thus reduced periods of credit.

- **Control of local circuits.**

The ability to energise and de-energise premises was included in the networks functionality, but the prospects for controlling individual circuits in a customer's premises might also be a functionality of use to a supplier. Under the Single Electricity Market suppliers will need to manage the risks of exposure to pool prices. Controlling customer load through agreed demand management arrangement could provide an economic method of managing an exposure to pool prices at times of high prices or in the event of an excursion in SMP.

- **Accuracy of billing.**

Improvements in the use of measured rather than estimated data, and the accuracy of that data, should substantially reduce the level of queries and the need to re-bill customers. Quantification of the reduced load on call centres and the consequent lower billing costs would need to be explored to assess the benefit that would accrue from this aspect.

#### VI. OTHER UTILITIES

- **Other Utilities.**

Once established a smart metering system would provide an obvious vehicle for extending the arrangements to other utilities. Gas metering would be an obvious addition but there may also be piggy-back applications in other areas such as the gas market and home security.

- **Energy efficiency.**

The role smart meters can play an important role in improving the efficiency of electricity use through changes in customer behaviour in response to price. However, the introduction of smart metering could provide routes to improved energy use through a messaging capability or measurement of the efficiency with which various applications (e.g. water heating) were achieved.

- **Reduced Emissions.**

Although usually construed as being capable of measuring energy use a smart meter system could display directly to customers the carbon footprint (CO<sub>2</sub> emissions) of various activities. This would be particularly the case if a system were designed to handle the measurement of both gas and electricity.

- **Capacity Margins.**

Demand side effects are generally thought to be of use to energy suppliers but producers could also make use of such arrangements to shave peak demands or as a substitute for operating peaking plant with high variable costs. The prospects for this might also be assessed.

- **Fuel poor and vulnerable customers.**

Smart metering could also provide information to social services that would enable the support of vulnerable customers, either through the provision of restricted supplies at times of financial hardship, or simply by monitoring the status of their supplies.

#### VII. ENERGY EFFICIENCY

It has been estimated that EU energy consumption is around 20% higher than can be justified on economic grounds. This has led to the view that there is a large potential for unrealised economic energy savings which can be realized through energy services and other end-use efficiency measures. In pursuit of this objective the European Commission adopted EU Directive EC 2006/32 on 5th April 2006. Article 13 of this Directive requires that where technically possible and prospectively economic, energy metering should record the time of use and customer billing should be sufficiently comprehensive so as to enable the self regulation of energy consumption. The Directive came into force on 17<sup>th</sup> May 2006 and had to be implemented by Member States by 17th May 2008. Smart metering is believed to be one technology whereby end-use energy efficiency can be encouraged, and thus emissions of greenhouse gases reduced [5].

as a measure to combat climate change. Therefore a lot of countries installed or would like install smart meters. The expected evolution in the electricity sector - smart meters

installation is illustrated in Figure 2 for the countries that provided historical and/or forecast data for a significant number of years [1].

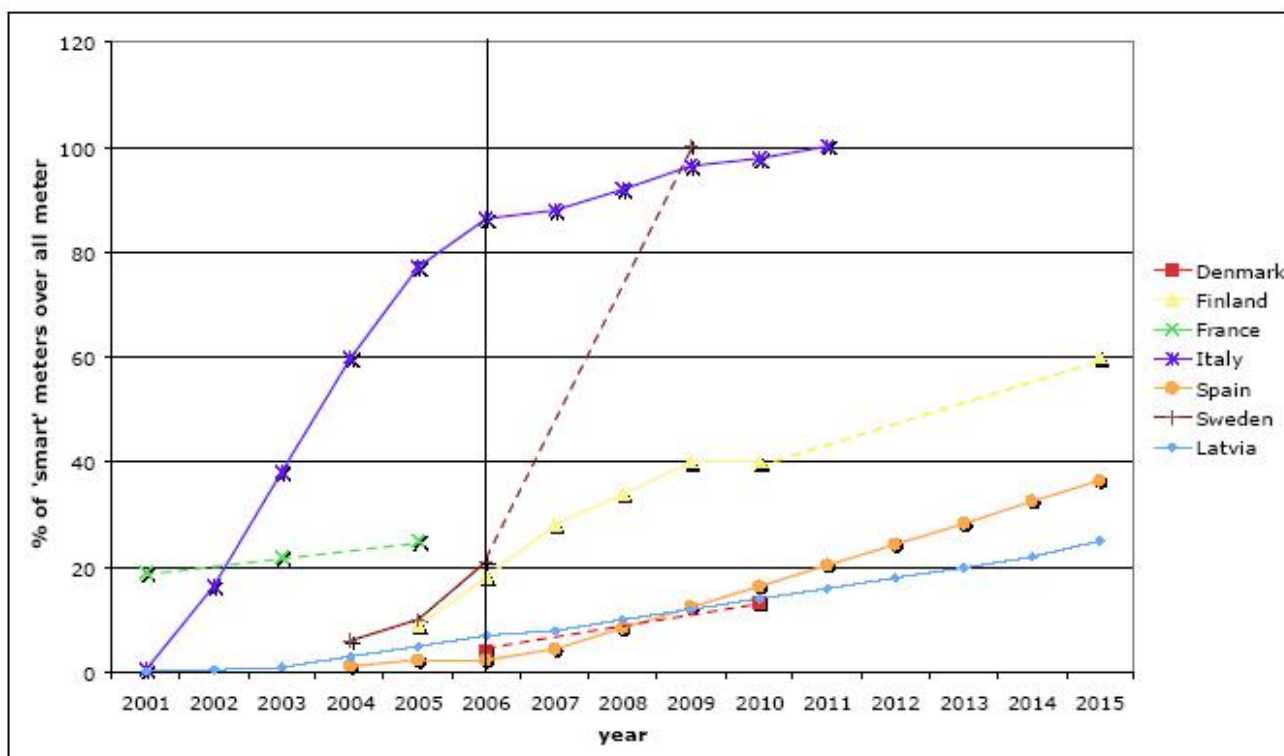


Figure 2: Smart meters installation: expected evolution in the electricity sector

### VIII. CONCLUSION

In conclusion, smart metering technologies are technically feasible and mature, at least for the electricity sector and at least on the technical side. Many manufacturers can supply competitive solutions, based on different functionalities, architecture and telecommunication systems. Smart metering programme has the potential to bring a wide range of prospective benefits to network operators (either as the owner of network assets or as a provider of metering system), supply businesses, customers, and government in the pursuit of wider policy objectives.

### ACKNOWLEDGMENT

This work was supported by Scientific Grant Agency of the Ministry of Education of Slovak Republic and the Slovak Academy of Sciences under the project VEGA No. 1/4072/07 and by the Slovak Research and Development Agency under the contract No. APVV-0385-07.

### REFERENCES

- [1] Smart metering with a Focus on Electricity Regulation, Ref: E07-RMF-04-03, 31 October 2007, Council of European Energy Regulators ASBL 28 rue le Titiens, 1000 Bruxelles
- [2] CER (2007). Demand side management and smart metering. Consultation paper, 13th March 2007. Available from: [www.cer.ie](http://www.cer.ie).
- [3] Ofgem (2006a). Domestic Meter Innovation. Consultation Document 20/06. Available from: [www.ofgem.gov.uk](http://www.ofgem.gov.uk).
- [4] EC (2005). Directive 2005/89/EC of the European Parliament and of the Council of 18 January 2006 concerning measures to safeguard security of electricity supply and infrastructure investment. Available from: [http://eur-lex.europa.eu/LexUriServ/site/en/oj/2006/l\\_033/l\\_03320060204en00220027.pdf](http://eur-lex.europa.eu/LexUriServ/site/en/oj/2006/l_033/l_03320060204en00220027.pdf).
- [5] EC (2006). Directive 2006/32/EC of the European Parliament and of the Council of 5 April 2006 on energy end-use efficiency and energy services and repealing Council Directive 93/76/EEC. Available from: [http://eur-lex.europa.eu/LexUriServ/site/en/oj/2006/l\\_114/l\\_11420060427en00640085.pdf](http://eur-lex.europa.eu/LexUriServ/site/en/oj/2006/l_114/l_11420060427en00640085.pdf).

# Numerical Solution of Induction Heating in 2D

Dušan MEDVEĎ

Department of Electric Power Engineering, FEI TU of Košice, Slovak Republic

Dusan.Medved@tuke.sk

**Abstract**—This paper deals with the numerical solution of induction heating in two dimensional geometries. There are analyzed electromagnetic and thermal field and their specifications.

**Keywords**—Induction heating, numerical modeling, finite element method.

## I. INTRODUCTION

The numerical modeling of induction heating of metal materials is given by two generally non-linear second-order differential partial equations of the parabolic type that describe the distribution of the electromagnetic and thermal fields. Coefficients of these equations commonly contain physical parameters of the heated material that are function of other quantities (temperature, frequency, pressure, etc.).

## II. MATHEMATICAL MODEL

Mathematical modeling is one of the major factors in the successful design of induction heating systems. Theoretical models may vary from a simple hand-calculated equation to a very complicated numerical analysis which can require several hours of computational work using computers. The choice of a particular theoretical model depends on several factors, including the complexity of the engineering problem, required accuracy, time limitations and cost.

### A. Electromagnetic Field

The technique of calculating electromagnetic field depends on the ability to solve Maxwell's equations. For general, time-varying electromagnetic fields, Maxwell's equations in differential form can be written as

$$\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} \quad (1)$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \quad (2)$$

$$\nabla \cdot \mathbf{B} = 0 \quad (3)$$

$$\nabla \cdot \mathbf{D} = \rho_0 \quad (4)$$

where  $\mathbf{E}$  is electric field intensity,  $\mathbf{D}$  is electric flux density,  $\mathbf{H}$  is magnetic field intensity,  $\mathbf{B}$  is magnetic flux density,  $\mathbf{J}$  is conduction current density and  $\rho_0$  is electric charge density.

The above-described Maxwell's equations are in indefinite form because the number of equation is less than the number of unknowns. These equations become definite when the relations between the field quantities are specified. The following constitutive relations are additional and hold true for a linear isotropic medium.

$$\mathbf{D} = \varepsilon_r \cdot \varepsilon_0 \cdot \mathbf{E} \quad (5)$$

$$\mathbf{B} = \mu_r \cdot \mu_0 \cdot \mathbf{H} \quad (6)$$

$$\mathbf{J} = \gamma \cdot \mathbf{E} \quad (7)$$

where the parameters  $\varepsilon_r$ ,  $\mu_r$ ,  $\gamma$  denote the relative permittivity, relative magnetic permeability and electrical conductivity of the material.

After some vector algebra and using equations (1), (2) and (6), it is possible to show that

$$\nabla \times \left( \frac{1}{\gamma} \cdot \nabla \times \mathbf{H} \right) = -\mu_r \cdot \mu_0 \cdot \frac{\partial \mathbf{H}}{\partial t} \quad (8)$$

$$\nabla \times \left( \frac{1}{\mu_r} \cdot \nabla \times \mathbf{E} \right) = -\gamma \cdot \varepsilon_0 \cdot \frac{\partial \mathbf{E}}{\partial t} \quad (9)$$

Since the magnetic flux density  $\mathbf{B}$  satisfied a zero divergence condition (3), it can be expressed in terms of a magnetic vector potential  $\mathbf{A}$  as

$$\mathbf{B} = \nabla \times \mathbf{A} \quad (10)$$

And then, from (2) and (10), it follows that

$$\nabla \times \mathbf{E} = -\nabla \times \frac{\partial \mathbf{A}}{\partial t} \quad (11)$$

Therefore, after integration, one can obtain

$$\mathbf{E} = -\frac{\partial \mathbf{A}}{\partial t} - \nabla \cdot \varphi \quad (12)$$

where  $\varphi$  is the electric scalar potential. Equation (7) can be written as

$$\mathbf{J} = -\gamma \cdot \frac{\partial \mathbf{A}}{\partial t} + \mathbf{J}_s \quad (13)$$

where  $\mathbf{J}_s = -\gamma \cdot \nabla \cdot \varphi$  is the source current density in the induction coil.

Taking the material properties as being piecewise continuous and neglecting the hysteresis and magnetic saturation it can be shown that

$$\frac{1}{\mu_r \cdot \mu_0} \cdot (\nabla \times \nabla \times \mathbf{A}) = \mathbf{J}_s - \gamma \cdot \frac{\partial \mathbf{A}}{\partial t} \quad (14)$$

For the great majority of induction heating application it is possible to further simplify the mathematical model by assuming that the currents have a steady-state quality. Therefore, with this assumption we can conclude that the electromagnetic field quantities in Maxwell's equations are harmonically oscillating functions with a single frequency. This field can be described by the following equations, which are derived after some vector algebra from (8), (9) and (14).

$$\frac{1}{\gamma} \cdot \nabla^2 \mathbf{H} = \mathbf{j} \cdot \omega \cdot \mu_r \cdot \mu_0 \cdot \mathbf{H} \quad (15)$$

$$\frac{1}{\mu_r} \cdot \nabla^2 \mathbf{E} = \mathbf{j} \cdot \omega \cdot \mu_r \cdot \mu_0 \cdot \mathbf{E} \quad (16)$$

$$\frac{1}{\mu_r \cdot \mu_0} \cdot \nabla^2 \mathbf{A} = -\mathbf{J}_s + \mathbf{j} \cdot \omega \cdot \gamma \cdot \mathbf{A} \quad (17)$$

where  $\nabla^2$  is the Laplacian, which has different forms in Cartesian and cylindrical coordinates.

Equations (15) and (16) are valid for general 2D and 3D fields and allow one to find all of the required design parameters of the induction system such as current, power, coil impedance and heat source density induced by eddy currents.

In some problems it is possible to reduce the 3D field to a combination of 2D form. The boundary of the region is selected such that the magnetic vector potential  $\mathbf{A}$  is zero along the boundary (Dirichlet condition) or its gradient is negligibly small along the boundary compared to its value elsewhere in the region (Neumann condition  $\frac{\partial \mathbf{A}}{\partial n} = 0$ ). Therefore, the heat transfer equation (see next section B) and equation (17), with their initial and boundary conditions, fully describe the electro-thermal process in a great majority of conventional cylindrical induction heat treatment systems.

### B. Thermal Field

In general, the transient (time-dependent) heat transfer process in a metal workpiece can be described by the Fourier equation:

$$c \cdot \rho \cdot \frac{\partial \vartheta}{\partial t} + \nabla \cdot (-\lambda \cdot \nabla \cdot \vartheta) = q \quad (18)$$

where  $\vartheta$  is temperature,  $\rho$  is the density of the metal,  $c$  is the specific heat,  $\lambda$  is the thermal conductivity of the metal and  $q$  is the heat source density induced by eddy currents per unit time in a unit volume (heat generation). This heat source density is obtained by solving the electromagnetic problem.

Equation (18), with suitable boundary and initial conditions, represents the 3D temperature distribution at any time and at any point in the workpiece. The initial temperature condition refers to the temperature profile within the workpiece at time  $t = 0$ . Therefore that condition is required only when dealing with a transient heat transfer problem where the temperature is a function not only of the space coordinates but also of time. The initial temperature distribution is usually uniform and corresponds to the ambient temperature.

For most engineering induction heating problems, boundary conditions combine the heat losses due to convection and radiation. In this case, the boundary condition can be expressed as

$$-\lambda \cdot \frac{\partial \vartheta}{\partial t} = \alpha \cdot (\vartheta_s - \vartheta_a) + C_s \cdot (\vartheta_s^4 - \vartheta_a^4) + Q_s \quad (19)$$

where  $\frac{\partial \vartheta}{\partial n}$  is the temperature gradient in a direction normal to the surface at the point under consideration,  $\alpha$  is the convection surface heat transfer coefficient,  $C_s$  is the radiation heat loss coefficient,  $Q_s$  is the surface loss and  $n$  denotes the normal to the boundary surface.

If the heated body is geometrically symmetrical along the axis of symmetry, the Neumann boundary condition can be formulated as

$$\frac{\partial \vartheta}{\partial n} = 0 \quad (20)$$

The Neumann boundary condition implies that the temperature gradient in a direction normal to the axis of symmetry is zero. In other words, there is no heat exchange at the axis of symmetry. This boundary condition can also be applied in the case of a perfectly insulated workpiece.

### III. NUMERICAL COMPUTATION OF THE HEATING PROCESS

The analytical methods and equivalent circuit coil design methods no longer satisfy the modern designer because of some restrictions. Rather than use simple computational techniques with many restrictions and poor accuracy, modern induction heating specialists are currently turning to highly effective numerical methods such as finite difference, finite element, mutual impedance and boundary element methods. These methods are widely and successfully used in the computation of electromagnetic and heat transfer problems.

#### A. Finite Difference Method

The finite difference method (FDM) has been used extensively for solving both heat transfer and electromagnetic problems. It is particularly easy to apply when the modeling area has cylindrical or rectangular shapes.

The computation procedure consists of replacing each partial derivative of the governing equations (17), (18) by a finite difference that couples the value of the unknown variable (i.e. temperature or magnetic vector potential) at an approximation node with its value in the surrounding area. This method provides a point-wise approximation of the partial differential equation and is quite universal because of its generality and its relative simplicity of application.

For two variables, the value of a variable at a node on the mesh can be expressed in terms of its neighboring values and separation distance (called a space step)  $h$  as in the expressions:

$$\frac{\partial \vartheta}{\partial x} = \frac{\vartheta_{i+1} - \vartheta_i}{h} + O(h) \quad (21)$$

$$\frac{\partial \vartheta}{\partial x} = \frac{\vartheta_i - \vartheta_{i-1}}{h} + O(h) \quad (22)$$

$$\frac{\partial \vartheta}{\partial x} = \frac{\vartheta_{i+1} - \vartheta_{i-1}}{2 \cdot h} + O(h) \quad (23)$$

$$\frac{\partial^2 \vartheta}{\partial x^2} = \frac{\vartheta_{i+1} - 2 \cdot \vartheta_i + \vartheta_{i-1}}{h^2} + O(h^2) \quad (24)$$

Where equation (21) denotes forward difference, (22) backward difference and (23) central difference method. The notation  $O(h)$  is used to show that the error involved in the approximation is on the order of  $h$ . Similarly,  $O(h^2)$  is for the approximation error on the order of  $h^2$ , which is more accurate than one on the order of  $h$ .

Substitution of the finite difference templates into the electromagnetic and heat transfer partial differential equations gives the local approximation. By assembling all local approximations and taking into account the proper initial and boundary conditions, one can obtain a set of simultaneous algebraic equations that can be solved with respect to unknown variables (i.e.  $\vartheta$ ,  $\mathbf{A}$ ,  $\mathbf{E}$ ,  $\mathbf{H}$  or  $\mathbf{B}$ ) at each node of the mesh. The solution can be obtained either by iterative techniques or by direct matrix inversion methods.

### B. Finite Element Method

This numerical technique (FEM) is very popular numerical tools and grand improvement in computer capabilities has boosted the development of several variations of the FEM.

As was mentioned in section A, the finite difference method provides a point-wise approximation and FEM provides an element-wise approximation of governing equations.

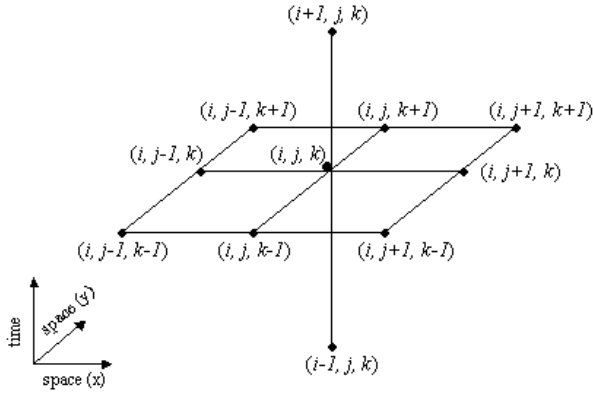


Fig. 1. Example of simplified 2D form for 2D approximation.

Since induction heating is a complex combination of electromagnetic and heat transfer phenomena, the FEM is suitable for solving both of them. The solution of electromagnetic field computation is typically obtained by minimizing the energy functional that corresponds to the governing equation (17) instead of solving that equation directly. The energy functional corresponding to the 2D governing equation (17) can be written in the form:

$$F = \int_V \frac{1}{2 \cdot \mu_r \cdot \mu_0} \cdot \left( \left| \frac{\partial A}{\partial x} \right|^2 + \left| \frac{\partial A}{\partial y} \right|^2 \right) \cdot dV + \quad (25)$$

$$\int_V \left( j \cdot \frac{\omega \cdot \gamma}{2} \cdot |A|^2 - J_s \cdot A \right) \cdot dV$$

where  $V$  is the total area of modeling and  $J_s$  is a source current density. The first terms inside the integrand represents energy of the magnetic field and second integral represents eddy currents and source current.

After some algebraic operations, the local matrix equation, which represents the minimization of the energy functional within any triangular element, can be written as

$$[[V]_e + j \cdot [W]_e] \cdot [A] = [Q]_e \quad (26)$$

where

$$[V]_e = \frac{1}{4 \cdot \mu_r \cdot \mu_0 \cdot \Delta} \cdot \begin{bmatrix} (b_l \cdot b_l + c_l \cdot c_l) & (b_l \cdot b_m + c_l \cdot c_m) & (b_l \cdot b_n + c_l \cdot c_n) \\ (b_m \cdot b_l + c_m \cdot c_l) & (b_m \cdot b_m + c_m \cdot c_m) & (b_m \cdot b_n + c_m \cdot c_n) \\ (b_n \cdot b_l + c_n \cdot c_l) & (b_n \cdot b_m + c_n \cdot c_m) & (b_n \cdot b_n + c_n \cdot c_n) \end{bmatrix} \quad (27)$$

$$\begin{bmatrix} a_l & a_m & a_n \\ b_l & b_m & b_n \\ c_l & c_m & c_n \end{bmatrix} = \begin{bmatrix} (x_m \cdot y_n - x_n \cdot y_m) & (x_m \cdot y_l - x_l \cdot y_n) & (x_l \cdot y_m - x_m \cdot y_l) \\ (y_m - y_l) & (y_n - y_l) & (y_l - y_m) \\ (x_n - x_m) & (x_l - x_n) & (x_m - x_l) \end{bmatrix} \quad (28)$$

$$[W]_e = \frac{\omega \cdot \gamma \cdot \Delta}{12} \cdot \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix} \quad (29)$$

$$[Q]_e = \frac{J_s \cdot \Delta}{3} \cdot \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \quad (30)$$

$$[A]_e = \begin{bmatrix} A_l \\ A_m \\ A_n \end{bmatrix} \quad (31)$$

where  $\Delta$  is a cross-sectional area of a particular triangular.

After assembling all local matrices of finite elements and specifying the corresponding boundary conditions, a global matrix equation can be obtained:

$$[G] \cdot [A] = [Q] \quad (32)$$

After solving the system of algebraic equations and obtaining the distributions of the magnetic vector potential in the modeling region, it is possible to find all of the required output parameters of the electromagnetic field.

The induced current density in the charge:

$$J_e = -j \cdot \omega \cdot \gamma \cdot A \quad (33)$$

The total current density in the charge:

$$J = J_s - j \cdot \omega \cdot \gamma \cdot A \quad (34)$$

The magnetic flux density components  $B_x$  and  $B_y$  can be calculated as:

$$\frac{\partial A}{\partial y} = -B_x; \quad \frac{\partial A}{\partial x} = -B_y \quad (35)$$

From (35) the flux density can be obtained as

$$B = \sqrt{B_x^2 + B_y^2} \quad (36)$$

Magnetic field intensity

$$H = \frac{B}{\mu_r \cdot \mu_0} \quad (37)$$

Electric field intensity

$$E = -j \cdot \omega \cdot A \quad (38)$$

From a vector potential solution it is possible to compute the other important quantities of the process such as stored energy, flux leakage, total power loss, and coil impedance.

As one can see, the described FEM requires using the current density of the source (induction coil) as the input parameter. This is often the case for induction hardening application. In the majority of induction heating applications prior to hot and warm forming when using multiturn coils it is necessary to have not a current density but the voltage of the coil as the input parameter.

Similarly the solution of a heat transfer phenomena consists of solution of equation (18). In simplified example it is possible to show solution of Laplace's or Poisson's equation using FEM.

Laplace's equation in terms of Cartesian coordinate system in 2D is in the form

$$a \cdot \frac{\partial \vartheta}{\partial t} = \frac{\partial^2 \vartheta}{\partial x^2} + \frac{\partial^2 \vartheta}{\partial y^2} \quad (39)$$

As an example for above mentioned FEM (Fig.2) one can see the result (Fig 3) of solving Laplace equation over the domain using program in MATLAB and 36 quadrilateral elements. The boundary conditions are also show in the figure 2.

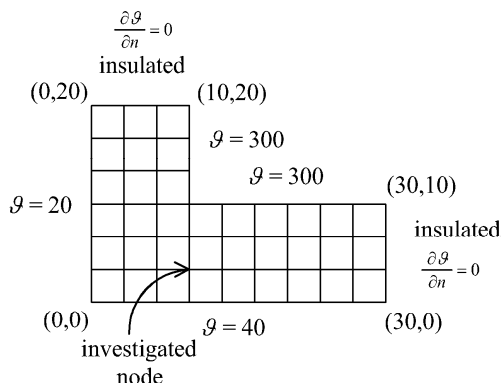


Fig. 2. Simplified 2D mesh for an example

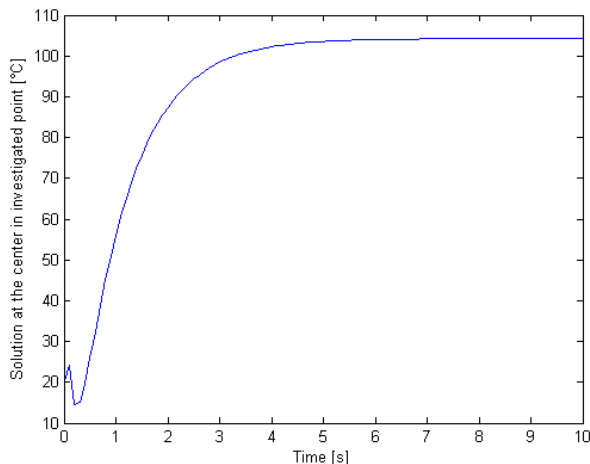


Fig. 3. Temperature development in the investigated point

#### IV. CONCLUSION

As one can see the solution of induction heating using FEM is quite simple. The accuracy of the numerical approximation of the governing partial differential equations improves with a finer mesh. In other words, the more finite elements used in the simulation, the better the approximation will be and the closer a numerical solution gets to the exact solution of the governing equations.

The presented method is also suitable for various shape of charge. The results also correspond to empirical experiences from solving of thermal and electromagnetic field.

#### ACKNOWLEDGMENT

This work was supported by project VEGA SR No. 1/4070/07.

#### REFERENCES

- [1] Kwon Y. W.: *The finite element method using MATLAB*. CRC-Press: 1997. 527 s. ISBN 0849396530.
- [2] Zienkiewicz, O. C., Taylor, R. L., Zhu, J. Z.: *The Finite Element Method: Its Basic and Fundamentals*. Butterworth Heinmann: 2005. ISBN 0-7506-6320-0.
- [3] Sadiku, M. N.: *Numerical Techniques in Electromagnetics*. CRC Press: 2000. 760 s. ISBN: 978084931
- [4] Won Young Yang: *Applied Numerical Methods Using MATLAB*. Wiley-Interscience: 2005. 528 s. ISBN 0471698334.3950.
- [5] Rudnev, V., Loveless, D., Cook, R., Black, M.: *Handbook of Induction Heating*. New York: Marcel Dekker, 2003. 777 pp. ISBN 0-8247-0848-2.
- [6] Kreith, F.: *CRC Handbook of Thermal Engineering*. CRC; 1. edition. 1999. ISBN 084939581X.
- [7] Medveď, D.: *Ohrev feromagnetických materiálov do Curieho teploty indukčnou metódou*. Dizertačná práca. Košice: Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky, 2008. 170 s.

# Performance evaluation of error resilience and error concealment in H.264

<sup>1</sup>Ján MOCHNÁČ, <sup>1</sup>Pavol KOCAN

<sup>1</sup>Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

<sup>1</sup>jan.mochnac@tuke.sk, <sup>1</sup>pavol.kocan@tuke.sk

**Abstract**—Video transmitted over noisy channels, e.g. wireless channels, is subject to packet losses, which degrade the visual quality. With growing interest in new and complex services implemented in the last generation of wireless networks also grows an interest in offering high quality video transmission. This paper is focused on performance evaluation of two different approaches which can be used to combat transmission errors, namely error resilience and error concealment, while H.264 video coding standard is used to compress video.

**Keywords**—Error concealment, error resilience, H.264, JM reference software.

## I. INTRODUCTION

Video transmitted over networks based on IP protocol is always subject to packet loss due to network congestion and channel noise. By using compressed video errors could propagate to the subsequent frames with resulting worse video quality. On the one hand, to combat transmission errors traditional error control and recovery schemes for data communication were adapted. These techniques introduced some redundancy which is helpful in lossless reconstruction of damaged video signal, but additional redundancy is also the main disadvantage while it is required bigger bitrate at the output of the encoder. On the other hand, signal reconstruction and error concealment have been proposed to obtain close approximation of the original signal or attempt to make the output signal at the decoder less objectionable to human eyes. Error concealment utilizes statistical redundancy which is always presented in video streams due to various reasons, e.g. coding delay, implementation complexity. It is employed in decoder and it has no dependence on sender [1]. Thus it utilizes only available video data. While in data transmission is required lossless delivery, in video transmission it is not necessary to deliver lossless video signal. Thus human eyes can tolerate a certain distortion in image and video signals.

After H.263 finalization in 1995, Video Coding Expert Group (VCEG) worked on new standard for low-bit rate video communication called H.26L. In 2001, VCEG and (MPEG) formed the Joint Video Team (JVT). JVT proceeded further with development of H.26L. In 2003, as a result of its effort were introduced two identical specifications, namely ITU-T H.264/AVC and ISO MPEG-4 Part 10. The new standard is officially entitled Advanced Video Coding (AVC). H.264 offers not only better coding efficiency but also focuses on error resilience and the adaptability to various networks. Error resilience and improved network adaptation is achieved through two-layered structure design. A video coding layer (VCL) is designed to obtain highly compressed video data.

Thus it includes core compression engine which is independent on networks. The network abstraction layer (NAL) formats VCL data and adds header information to various transportation protocols and storage media.

## II. ERROR RESILIENCE IN H.264

As well as MPEG-4 and other previous video coding standards, H.264 offers several error resilience tools. Some of them were already introduced in the prior standards others are implemented in innovated way [2], [3].

### A. Intra placement

Intra placement on the macroblock, slice or picture level is used primarily to combat drifting effects.

- H.264 allows intra macroblock prediction even from Inter macroblocks. It allows achieve better coding efficiency.
- Two forms of slices, which contain only intra macroblocks: Intra slices and Instantaneous Decoder Refresh (IDR) slices. IDR slices always form IDR picture. IDR picture has a stronger resynchronization property than a picture that contains only intra slices. After reception of these slices the decoder deletes all its previous stored reference pictures for resynchronization [3].

### B. Picture segmentation

If picture segmentation (PS) is used, a video picture is coded as one or more slices, each containing an integral number of macroblocks in a picture. The number of macroblocks per slice need not be constant within a picture. Unless flexible macroblock ordering is not used, macroblocks are ordered in raster scan order in the slices. Every macroblock in a picture must be coded in exactly one slice [3].

### C. Data partitioning

Data partitioning (DP) creates more bit strings, called partitions, per slice and allocates all symbols of a slice into an individual partition that have a close semantic relationship with each other. The following partitioning is defined in the specification:

- partition A contains slice header, macroblock types, quantization parameters, prediction modes and motion vectors
- partition B contains residual information of intracoded macroblocks
- partition C contains residual information of intercoded macroblocks

The idea behind DP is still to be able to use information from correctly received partitions when one of the partitions is lost [3].



#### D. Flexible macroblock ordering

Flexible macroblock ordering (FMO) allows assigning MBs to slices in an order other than the scan order. The main goal of this technique is to scatter possible errors to the whole frame as equally as possible. It is done by choosing the slice groups in such a way that no macroblock and its neighbors belong to the same group. When an error occurs (slice is lost), it will be easy to recover it in this case. Every macroblock of a picture can be assigned to an arbitrary slice group. The only limitation is that all MBs of every slice group are coded in raster scan order [3].

### III. ERROR CONCEALMENT IN H.264

It is assumed that in case of error slices, these are not decoded but discarded before decoding. Correctly received slices are decoded first and the lost slices are concealed thereafter. Information about correctly received MB or about lost MB is recorded in a macroblock based status map of the frame.

Spatial error concealment (EC) in H.264/AVC is performed by computing weighted pixel averaging. Every weight is calculated as inverse distance between the pixel to be concealed and the boundary pixels.

For temporal EC is chosen that of MVs which minimize boundary match distortion among the set of candidate of MVs including zero MV. Thus, by using chosen MV it is achieved smallest luminance change across block boundaries when the block is inserted into its place [4], [5].

### IV. PERFORMANCE EVALUATION

We have used JM reference software in our simulation. Two different sequences have been compressed without any error resilience and subsequently different error resilience tools have been applied. IDRs were used every 12 pictures and FMO pattern was raster scan. We have investigated loss rate from 0% to 10%. Damaged video sequences have been imported to MATLAB for PSNR evaluation.

As we can see in the Table I and also in the Table II, for low loss rate it is possible to use no error resilience or error concealment method, with increasing amount of lost packets this tools become necessary. As we can see the performance of error resilience tools is better than chosen error concealment method, but EC is almost 3dB better than no error resilience for 10% packet loss rate.

As we can see in this two tables, it also clear that performance of error resilience and error concealment methods is strongly dependent on video content. Mobile sequence contains much more motion than Foreman sequence with resulting lower PSNR.

TABLE I  
PSNR FOR ERROR RESILIENCE AND ERROR CONCEALMENT, FOREMAN SEQUENCE

Loss rate [%]	No error resilience	DP	FMO	IDR	PS	EC
0	39.00	38.95	39.01	38.98	38.96	38.99
1	38.89	38.88	38.93	38.88	38.85	38.75
2	38.12	38.79	38.85	38.81	38.84	38.71
5	36.52	38.01	38.16	37.99	38.01	37.52
10	31.99	36.22	36.45	36.23	36.32	34.93

TABLE II  
PSNR FOR ERROR RESILIENCE AND ERROR CONCEALMENT, MOBILE SEQUENCE

Loss rate [%]	No error resilience	DP	FMO	IDR	PS	EC
0	37.98	37.99	37.99	37.95	37.97	37.97
1	37.91	37.97	37.97	37.95	37.93	37.93
2	36.99	37.79	37.80	37.68	37.64	37.35
5	34.12	36.12	36.14	36.01	36.01	35.69
10	26.45	32.16	32.25	32.15	32.20	30.93

It is also interesting to compare output bit rate at the encoder without and with error resilience tools (Table III). DP requires only small increase in bit rate, but it gives very good results in PSNR evaluation, also increase in bit rate for FMO is acceptable for most application. On the other hand placing IDR pictures to the bit stream every 12 pictures is not very effective, with good resilience property, but with considerably high increase in output bit rate.

TABLE III  
OUTPUT BIT RATE AT THE H.264 ENCODER FOR FOREMAN SEQUENCE

	No error resilience	DP	FMO	IDR	PS	EC
Output bit rate [kbit/s]	26.267	26.93	32.53	49.24	28.07	67.81

### V. CONCLUSION

In this paper we have evaluated performance of error resilience tools and error concealment method used in H.264 reference software. As we have shown, both approaches leads to the significant better results than without them. Difference between error resilience and error concealment is small for video sequences with simple content like Foreman. Effectiveness of error resilience tools can be viewed in video sequences with a lot of motion like Mobile.

The best way to achieve the better visual quality is combination of error resilience with error concealment. But the main disadvantage of error resilience is higher output bitrate, while for error concealment less visual quality is the main disadvantage.

### ACKNOWLEDGMENT

The work presented in this paper was supported by Grant of Ministry of Education and Academy of Science of Slovak Republic VEGA under Grand No 1/4088/07.

### REFERENCES

- [1] Y. WANG and Q.-F. ZHU, "Error Control and Concealment for Video Communication: A Review," *Proceedings of the IEEE*, vol. 86, no. 5, pp. 974-998, May 1998.
- [2] S. KUMAR, L. XU, M. MANDAL, and S. PANCHANATHAN, "Error Resiliency Schemes in H.264/AVC Standard," *Elsevier J. of Visual Communication and Image Representation*, vol. 17(2), April 2006.
- [3] S. M. Jn MOCHNÁČ, "Error resilience tools in the MPEG-4 and H.264 video coding standards," *In Proceedings of Radioelektronika 2008*, pp. 187-190, April 2008.
- [4] —, "Error Concealment Scheme Implemented in H.264/AVC," *In: ELMAR-2008 : Proceedings : 50th International Symposium*, vol. 1.
- [5] J. Polec, T. Karlubíková, S. Ondrušová, and K. Kotuliaková, "New Objective Criterion for Error Evaluation of Video Object Contour," *In: IWSSIP 2008. 15th International Workshop on Systems, Signals and Image Processing*.

# Automated oscilloscope measuring

Ján MOLNÁR

Dept. of Theoretical Electrical Engineering and Electrical measurement, FEI TU of Košice, Slovak Republic

jan.molnar@tuke.sk

**Abstract** — In the paper, communication between TEKTRONIX oscilloscope and computer via RS-232, consequently measurement via Internet are described. The evaluation version of Control Web program has been utilized to allow a remote measurement. Evaluated program receives the data from an oscilloscope in addition the received data saves to the file and distributes it to remote computer via Internet. Remote computer shows the shape of signals saved by oscilloscope using an external program. The External program is made in C++ language using C++ Builder software. The program purpose is to draw and process the measured waveform.

**Keywords** — Oscilloscope, Internet, RS-232, Control Web, Measurement.

## I. CONTROL WEB

The Control Web software is universal tool for development and visualization of control applications, data acquisition applications, saving and data evaluation and applications of interface human-machine. Definitely, this software respects existing standards for user interface, data interchange, data access, communication in computer network and hardware interface for data acquisition and data control.

The Control Web works in operating system environment with implemented software interface Win32 and supports many of industry standards. The ASCII drivers have been used for serial communication as the part of Control Web. The Control Web contains HTTP server to allow for network communicating. The server is integrated to the system. It means that it offers static files also dynamically generated files based on being application. Furthermore, the data can be received from HTML data formulary and a virtual measurement set up can be activated depending on client's requests. As to applications, there is possibility to communicate with each other.

## II. OSCILLOSCOPE TEKTRONIX

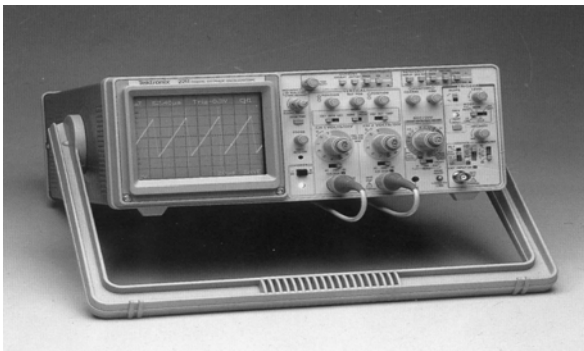
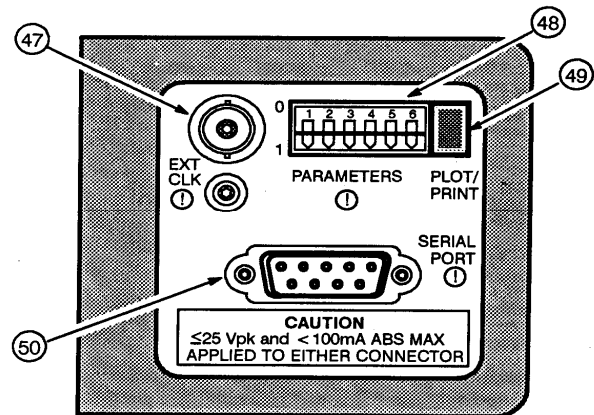


Fig. 1. The 2211 Oscilloscope

TEKTRONIX 2211 is portable analog - to - digital oscilloscope. It includes two input channels. In analog mode, the frequency can range from up to 50 MHz. In digital mode, it is from up to 1 MHz.

### 2.1 Description of oscilloscope interface RS-232

In the Fig. 2. is showed side panel of the oscilloscope, which is used for serial communication and for setting the communication parameters.



- 47. Input for external timing signal
- 48. Switch for setting of serial port
- 49. Press button for launching of data transfer from oscilloscope to the PC
- 50. Standard connector RS 232

Fig. 2. Side panel of the oscilloscope

This oscilloscope uses 9-pin DTE connector for serial communication. It is asynchronous transfer with transfer rate 300, 1200, 4800 or 9600 bps. Data string contains 8 data bits, start bit, stop bit and no parity bit.

## III. DECODING DATA

Transferred data from oscilloscope to the file are in HP-GL language (Hewlett – Packard Graphic Language). These data consist of complete information about signal curve and description of oscilloscope parameters setting (when signal is saving to the reference memory). The data represent the signal as a vector and alphanumeric values from oscilloscope display.

### 3.1 Examples of HP-GL commands

SP n; /Select pen

SC Xmin, X max, Ymin, Ymax; /Scale

PA x, y;	/Plot absolute
PU	/Pen Up
DF	/Set Defaults
SR	/Set character size
PD	/Pen Down
LB text	/label

We can determine next oscilloscope settings from acquired data: time base, trigger, number of scanned channels, voltage range, signal mode (AC/DC), and curve image. Following figure shows example of saved data

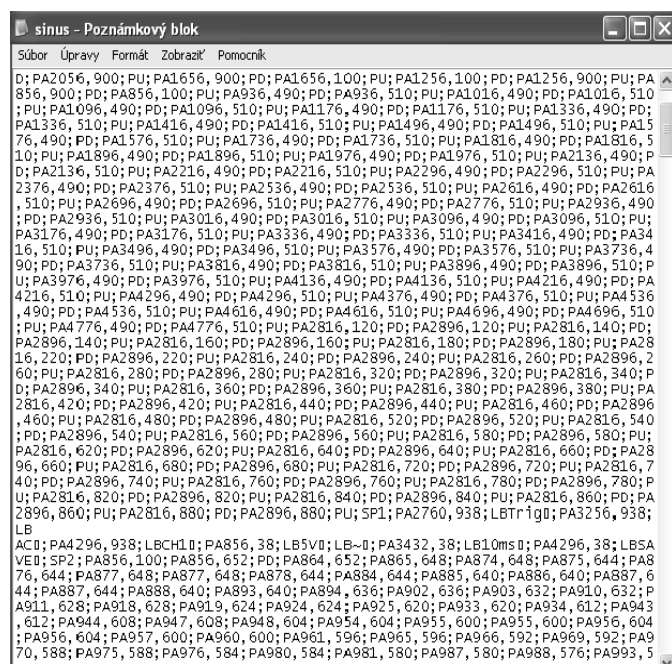


Fig. 3. Example of saved data

#### IV. PROGRAM DESCRIPTION

##### 4.1 Description of communication

Activating the application, the window named Oscilloscope local is shown at computer monitor, which is connected to the oscilloscope. In the same moment, the window named Oscilloscope remote at the monitor of remote computer is shown. Oscilloscope local and Oscilloscope - remote are mutual connected applications made by using Control Web that communicate via protocols TCP/IP. The measured data are transported via serial link to both local and remote PC, where are saved to text file. At moment when the transport has been finished, the data can be mutually processed in local and remote PC.

##### 4.2 Program environment

The same user interface is used for local and remote PC. Therefore, the program description is valid for local also for

remote computer. Basic panel window of described program is shown in the Fig 3.

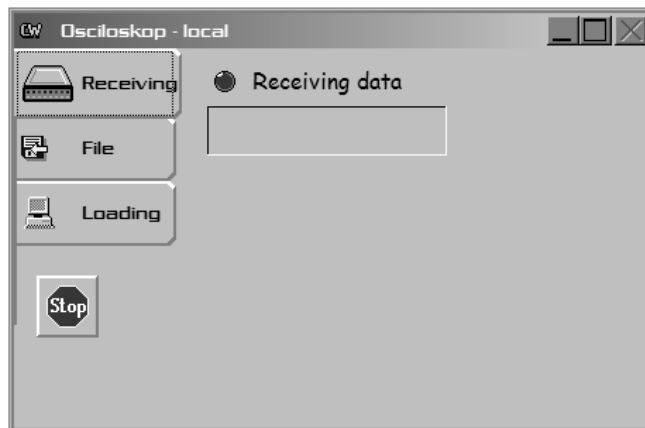


Fig. 3. The basic panel window

The basic panel window consists of 4 buttons: Receiving, File, Loading and Stop. After clicking on the button Receiving, the program waits for data from oscilloscope. The data could be transmitted, if the button 49 has to be hand pressed at the back panel of oscilloscope (see Fig. 2). After clicking on the button File, we can insert file name and URL.

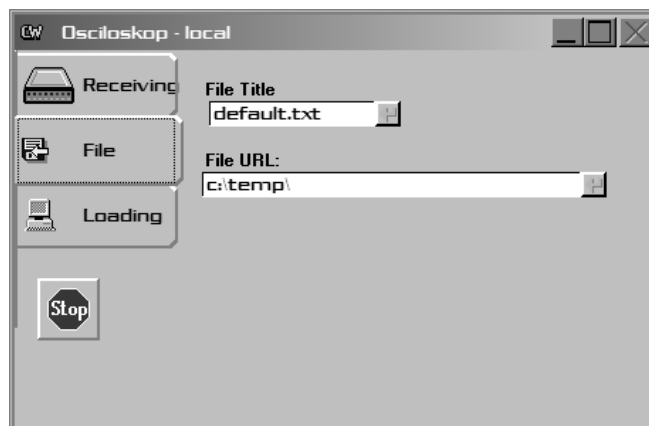


Fig. 4. Inserting file name and URL

After clicking on the button Loading, the program loads a saved file and consequently displays loaded curve by external program. In the cell "File Title " is necessary to insert the file name, where data will be saved.

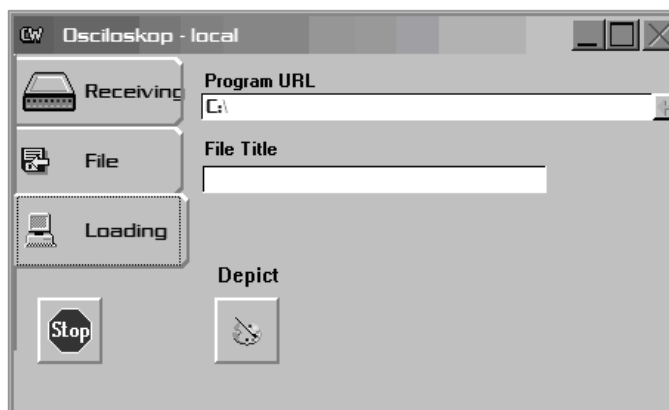


Fig. 5. Loading of saved file

By button Stop the programs in local and remote PC are finished. Appearance of the measured curve by oscilloscope is shown in the next figure.

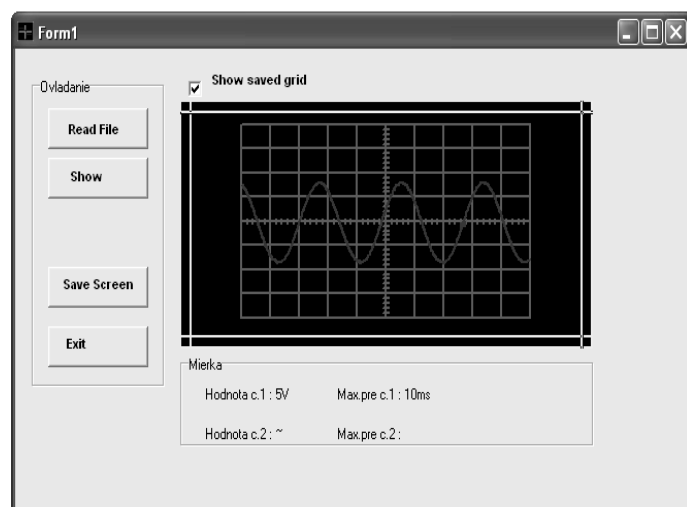


Fig. 6. Drawing of the measured curve

After click on “Read File” button, dialog window appears for selecting the File with stored data from oscilloscope. Then “Show” Button click shows the stored waveform we have saved before. The saved file is text type file. There is possibility to show also grid, which has been saved together with the data from oscilloscope. The program does not generate the grid, but it is saved in original form from oscilloscope data. For grid displaying there is check box labeled “Show saved grid”.

“Save Screen” screen button allows to save the screen – the waveform as a bitmap. After click the Save Screen button user just specifies the path and name to the new bitmap picture. It is not more in HP-GL language therefore it can be used in all programs working with bitmap graphics. Following figure (Fig. 7) shows such a saved waveform.

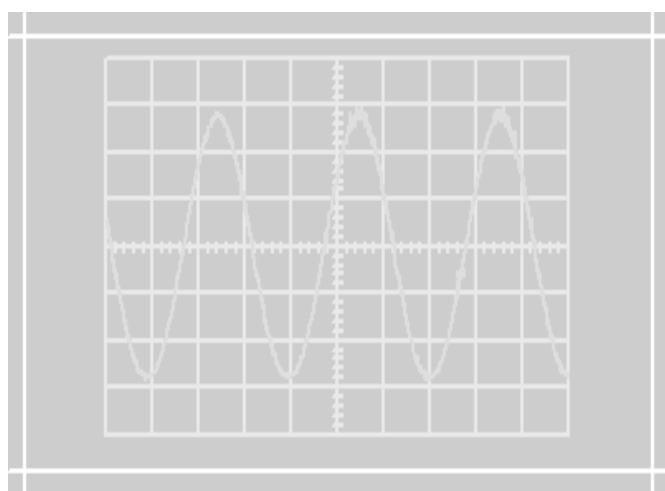


Fig. 7. Drawing of the measured curve

## V. CONCLUSION

Receiving the measured data, their distributing to remote oscilloscope and saving after finishing are main features of the described program. The curve drawing depends on the

external program. External program is able to draw and save measured curve. The curve is possible to use for follow-up postprocessing. From the waveform we are able to determine  $U_{\max}$ ,  $U_{ef}$ ,  $U_{str}$  and frequency.

## ACKNOWLEDGMENT

The paper has been prepared by the support of Slovak grant projects VEGA No. 1/4174/07, VEGA No. 1/0660/08, KEGA3/5227/07, KEGA 3/6386/08 and KEGA 3/6388/08.

## REFERENCES

- [1] TEKTRONIX: 2211 Oscilloscope operators' manual. 1988, Revised FEB 1989, Revised November 1989.
- [2] TEKTRONIX: GRABBER II Waveform transfer software for the TEK 2201/2211/2214. 1<sup>st</sup> edition, 1989.
- [3] Control Web 2000 - Průvodce systémem pro tvorbu a nasazení aplikací reálného času, Praha, Computer Press, 1999.
- [4] Kováč, D., Kováčová, I., Molnár, J.: Elektromagnetická Kompatibilita-meranie, TU Košice, ISBN 978-80-553-0151-8.
- [5] Kováč, D., Kováčová, I., Vince, T.: Elektromagnetická Kompatibilita, TU Košice, ISBN 978-80-553-0150-1.
- [6] Tomčíková, I., Špaldonová, D.: Elastomagnetic Sensor Field Determination using MATLAB. In: Acta elektrotechnica et Informatica. vol. 7, no. 3 (2007), p. 74-79 ISSN 1335-8243.
- [7] Tomčík, J., Tomčíková, I.: IT bezpečnosť automatizačných a SCADA systémov (1). In: AT&P Journal, roč. 13, č. 4(2006), s. 50 – 54. ISSN 1336-5010.
- [8] Kováčová, I., Kováč, D.: Modelling and Measuring of Electronic Circuits, textbook FEI TU Košice, ELFA s.r.o. Publisher, 1996, 92 pages, ISBN 80-88786-44-4.
- [9] Kováčová, I., Kováč, D.: EMC Compatibility of Power Semiconductor Converters and Inverters, Acta Electrotechnica et Informatica, 2003, No.2, Vol.3, pp.12-14.
- [10] Molnár, J., Kováčová, I.: Distance remote measurement of magnetic field, Acta Electrotechnica et Informatica, 2007, No.4, Vol.7, pp. 52-55, ISSN 1335-8243.
- [11] Molnár, J.: Remote measurement system in automobile scheme. In: 8<sup>th</sup> International PhD Workshop OWD 2006, Warszawa: Komitet badania naukowe, 2006. p. 400-403. ISBN 83-922242-1-3.
- [12] Molnár, J.: Magnetic field measurement via Internet. In: AMTEE '05, Seventh international conference on Advanced Methods in the Theory of Electrical Engineering, Plzeň, University of West Bohemia, 2005. pp. d59-d73. ISBN 80-7043-392-2.

# Electricity System of Libya and its Future

Ing. Maher NASR

Department of Electric Power Engineering, FEI TU of Košice, Slovak Republic

maher.nasr@tuke.sk

**Abstract**—This article deals with electricity system of Libya. The current transmission lines are very loaded and there is some new calculation for building of new ones. The main energy strategy is spreading of interconnection between other countries. In this article will be presented also what is the expected peak load in the future and the solution for ensure the sufficient amount of electric energy.

**Keywords**—Electricity system, transmission lines, energy of Libya.

## I. INTRODUCTION

Today, the electricity system of Libya is a state owned vertically structured power utility company and is responsible for generation, transmission and distribution of electric energy. The installed generation capacity was around 6612 MW while the peak load was 4422 MW in 2007. The transmission system consists mainly of 220 kV lines, but the expected load will increase so there are some new calculations, that you can see in the next chapter.

## II. THE ACTUAL ELECTRICITY SYSTEM OF LIBYA

The actual electricity system of Libya is controlled by state owned company GECOL. This company operates 30 electricity generation plants, mainly steam and simple-cycle gas-turbine units and diesel generators in rural areas. The company is also the sixth largest operator of water desalination plants in the world. More than \$1,000 million will be spent on 11 new desalination units over the next ten years consist mainly from based mainly on oil.

The overall electricity statistics in 2007 was as follows:

- Installed generation capacity: 6284 MW
- Generated energy: 25514 GWh
- Peak demand: 4422 MW
- 400 kV transmission system: 442 km
- 220 kV transmission system: 13631 km
- 33 and 66 kV sub-transmission system: 22258 km
- 11 kV distribution system: 50000 km
- Number of customers: 1190940
- Average consumption per 1 customer: 4,158 kWh
- Installed capacity of desalination system: 46514 m<sup>3</sup>/day.

The total energy that plants produced during the year 2007 was 25,415 TWh, what is the energy rate growth of 6,4 % compared to 2006. The annual electricity generation according to years 2000 to 2007 shows the next table.

TABLE I  
THE ANNUAL ELECTRICITY GENERATION

Year	2000	2001	2002	2003	2004	2005	2006	2007
Generation [TWh]	15,496	16,111	17,531	18,943	20,202	22,450	23,992	25,415

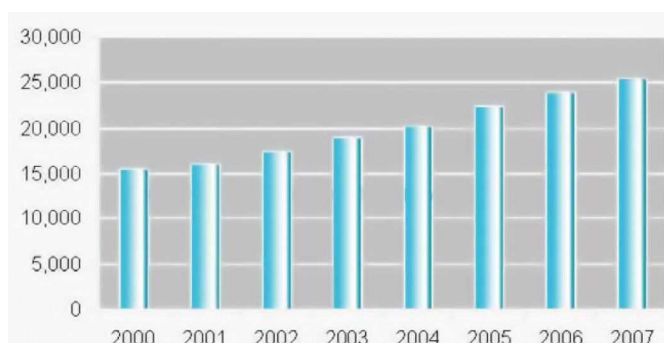


Fig. 1. Graphical representation of annual electricity generation

This generation increased every year for increased electricity demands. It was followed by spreading of electricity transmission lines.

The total portion of electricity transmission lines according to years 2000 to 2007 shows the next table.

TABLE II  
THE ANNUAL PORTION OF TRANSMISSION LINES

	400	220	66	30
2000	–	60	152	267
2001	–	62	156	268
2002	–	62	163	277
2003	–	62	166	281
2004	–	62	167	286
2005	2	64	169	302
2006	2	70	169	321
2007	3	70	175	355

The load of the network during the year 2007 is 4,420 GW, compared with the year 2006 was 4,005 GW, which was a growth rate of 10 %. The other loads are interpreted in the next table.

TABLE III  
THE ANNUAL ELECTRICITY LOAD

Year	2000	2001	2002	2003	2004	2005	2006	2007
Generation [GW]	2,630	2,934	3,081	3,341	3,612	3,857	4,005	4,420

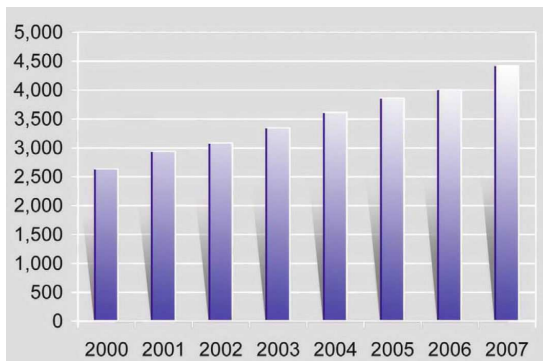


Fig. 2. Graphical representation of annual electricity load

The energy statistics of produced energy during the year 2007 according to type of generation is presented in following table.

TABLE IV  
THE ENERGY STATISTICS IN 2007

Energy produced [GWh]	Type of Generation
7713,335	Steam
12662,492	Nature gas
5138,255	Double cycle
25415,082	Total

The total number of substations shows the next table.

TABLE V  
THE PORTION OF SUBSTATIONS (TRANSFORMER CAPACITY)

[kV]	400	220	66	30
Number of stations	3	70	175	355
[MVA]	2400	13058	3559	9980

The total length of transmission lines shows the next table.

TABLE VI  
TRANSMISSION LINES

Item	1993	2000	2007
400 kV lines [km]	–	–	442
220 kV lines [km]	12640	12887	13723
66 kV lines [km]	10568	12475	13340
30 kV lines [km]	6166	6986	9410
Total [km]	29374	32348	36915

The electricity generation by fuel type shows the next graph.

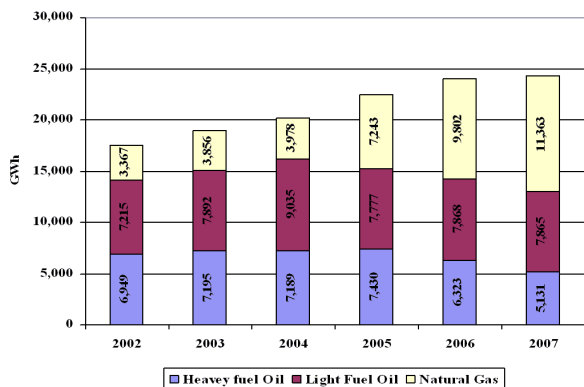


Fig. 3. Graphical representation of electricity generation by fuel type

The electricity consumption by customer type in 2007 shows the next graph.

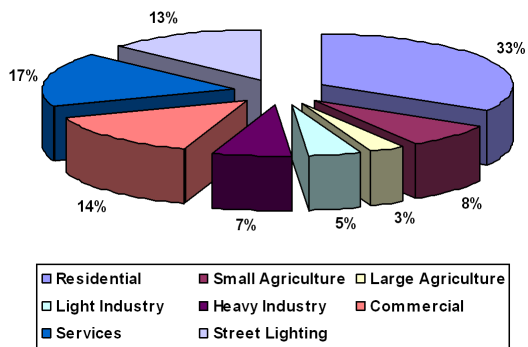


Fig. 4. Graphical representation of electricity consumption by customers

### III. THE FUTURE PLANS OF ELECTRICITY SYSTEM OF LIBYA

The main master plan of electricity system of Libya concludes objectives:

- securing and guarantee the electrical power supply to growing demand of electrical energy to all sectors in the country,
- increasing the level of security and adequacy of supply,
- reducing cost by improving service quality and efficiency,
- reducing technical and non technical losses,
- reinforcing international interconnections.

The long term load forecast:

- the peak load of electrical power in Libya is continuously increasing with a relatively high growth rate 8 % per annum,
- recent studies have shown that the expected peak demand in Libya in the year 2008 will be about 4670 MW and the figure is expected to reach 5450 MW by year 2010 and approximately 8000 MW in 2015.

TABLE VII  
THE LONG TERM LOAD FORECAST (IN MW)

Year	2008	2009	2010	2011	2012	2013	2014	2015
Total	4671	5045	5458	5884	6355	6863	7412	8005

These loads will require building new power capacities:

TABLE VIII  
NEW PROJECTS (IN MW)

Year	2008	2009	2010	2011	2012	2013	2014	2015
Total	164	1937	3386	4192	5047	5669	5779	5889

TABLE IX  
THE TOTAL EXPECTED PEAK LOAD (IN MW)

Year	2008	2009	2010	2011	2012	2013	2014	2015
Total	4835	6982	8834	10076	11402	12532	13191	13894

There is around 11000 MW generation capacity needed to be added during the period 2005 – 2015. Nearly 4000 MW are needed by the horizon 2010 with a mixed generation options

based on latest technology (steam and combined cycle) using natural gas. Also, under construction there is approximately 1500 MW of new power plants (750 MW in Benghazi C. C. and 750 MW in Misurata C. C). The other awarded contracts for building of new power plants are in West Mountain Ext. (312 MW), Zwitina gas plant (500 MW), Srir west gas plant (750 MW), Gulf steam plant (1400 MW), Tripoli West steam plant (1400 MW) and under contract is Sebha gas plant (750 MW).



Fig. 5. Planned generation projects

During the horizon 2010 the main infrastructure for a strong, reliable, flexible and adequate backbone of the new 400 kV grid will be executed. Therefore, the future plan of the Libyan transmission system is concentrated on 400 kV grids. There are 2 400/220 kV substations and 442 km of 400 kV lines now. Under construction are 7 substations and 2720 km of new 400 kV transmission lines.



Fig. 6. The new Libyan 400 kV network

The future plan of the Libyan transmission system expansion includes also the infrastructure of:

- 256 km (1000 km cables core length) of 220 kV cables in cities, i.e. Tripoli, Benghazi, Zawia and Musrata,
- 71 substations of 220 kV in different locations in the country,
- around 2000 km of new 220 kV of transmission lines.

IV. ELECTRICAL INTERCONNECTIONS OF LIBYA

The internal interconnection of the Libyan transmission system has been achieved since 1993 creating the high voltage 220 kV Libyan national grid. The electrical networks of Libya and Egypt are interconnected on 220 kV voltage level since May 1998. By the end of the last year 2008 the electrical interconnection between Libya and Tunisia is expected to be synchronized on the 220 kV.

TABLE X  
CONTRACTED ENERGY EXCHANGE (IN MWH)

	Libya – Egypt		Libya – Tunisia	
	Import	Export	Import	Export
2003	16708	–	25038	–
2004	8565	–	32184	–
2005	11164	504	32491	–
2006	30292	175	–	–
2007	6612	1875	–	–
Total	73341	2554	89713	0

It is expected to construct and operate the 500-400 kV link with Egypt by the year 2010 and with Tunisia 400 kV during the period 2010-2015. Under study is the electrical interconnection on 400 kV between Libya and Algeria and the submarine 400 kV cable between Libya and Italy.

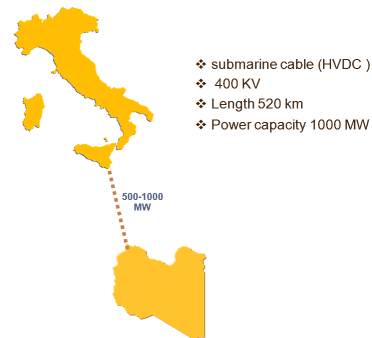


Fig. 7. Possible Interconnection with Europe: Libya – Italy (HVDC link)

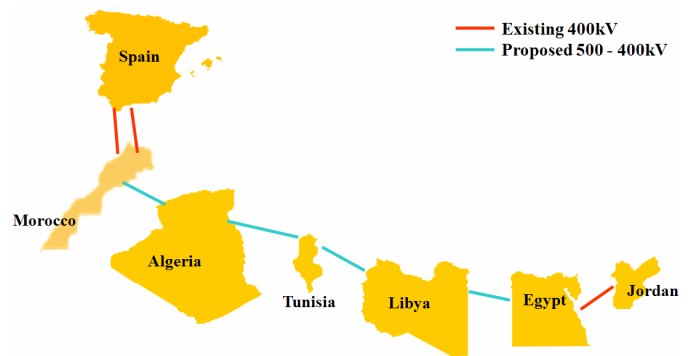


Fig. 8. Proposed 500-400 kV interconnection between ELTAM countries

V. CONCLUSION

Completion of 400kV network permits power wheeling of about 500 MW in both directions. Availability of large quantities of natural gas will lead to development of new power station, by which we can export electricity to the neighboring countries.

ACKNOWLEDGMENT

I would like to thank prof. Ing. Michal Kolcun, PhD. for supervising my study.

REFERENCES

[1] The Euro-Mediterranean Energy Partnership. [online] [cited 12.3.2009] <http://www.auptde.org/NewSite/UploadFiles/Activityfile/153.ppt>  
 [2] General Electricity Company of Libya. [online] [cited 12.3.2009] <http://www.gecol.ly/gecol/index.php>

# Adaptation of acoustic models for robust speech recognition

<sup>1</sup>Marek PAPCO, <sup>2</sup>Martin LOJKA

Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

<sup>1</sup>marek.papco@tuke.sk, <sup>2</sup>martin.lojka@tuke.sk

**Abstract**—This paper presents an adaptation of acoustic models for telephone characteristics in speech recognition. Our approach is to adapt acoustic models of high-quality speech to speech influenced by telephone line. The general framework for this adaptation with small amount of data is HTK toolkit. Several techniques for the adaptation of acoustic models were used to adapt acoustic models trained on Speechdat and Mobildat speech database. These techniques are MAP (Maximum a posteriori), STC (Semi-tied transform) and HLDA (heteroscedastic linear discriminant analysis). Adaptation data originate from IRKR system logs, that contain voice oriented dialogue between humans and system for providing weather forecast and traffic guide in telecommunication networks. HTK toolkit supports supervised and unsupervised adaptation. The article describes steps to successful adapt acoustic models with supervised offline adaptation technique. Preliminary experimental results of adapted models are compared with baseline acoustic models with no adaptation technique applied.

**Keywords**—acoustic model, adaptation, MAP, STC, HLDA

## I. INTRODUCTION

Acoustic mismatches between training and testing conditions of an automatic speech recognizer can significantly downgrade the recognition accuracy. ASR (automatic speech recognition) system obtain high recognition accuracy under laboratory environment, whereas its performance degrades under noisy environment. The main reasons of degradation are background noise, channel effects and reverberation. A wide variety of methods [1] have been proposed to deal with these problems. The simplest way for compensation is to modify the training data to be representative of the application environment and retrain acoustic model. But this approach is problematic in large-vocabulary recognition systems due to insufficiency of training data. Instead of retraining acoustic models, algorithms as MAP, MLLR, STC and HLDA are used to adapt parameters of acoustic model to new environment with only small amount of adaptation data.

### A. MAP

MAP [1], [2] or Bayesian estimation can effectively deal with data-sparse problems, as we can take advantage of prior information about existing models. We can adjust parameters of retrained models in such way, that limited new training data modify parameters of acoustic model guided by prior knowledge to compensate for adverse effect of a mismatch. For MAP adaptation purposes the informative priors that are used are the speaker independent model parameters. The update formula for state  $j$  and mixture component  $m$  is

$$\hat{\mu}_{jm} = \frac{N_{jm}}{N_{jm} + \tau} \bar{\mu}_{jm} + \frac{\tau}{N_{jm} + \tau} \mu_{jm} \quad (1)$$

where  $\tau$  is a weighting factor of the a priori knowledge to the adaptation data and  $N$  is the occupation likelihood of adaptation data, defined as,

$$N_{jm} = \sum_{r=1}^R \sum_{t=1}^{T_r} L_{jm}^r(t) \quad (2)$$

where  $\mu_{jm}$  is the mean of initial model and  $\bar{\mu}_{jm}$  is the mean of the observed adaptation data, defined as,

$$\bar{\mu}_{jm} = \frac{\sum_{r=1}^R \sum_{t=1}^{T_r} L_{jm}^r(t) \mathbf{o}_t^r}{\sum_{r=1}^R \sum_{t=1}^{T_r} L_{jm}^r(t)} \quad (3)$$

where  $\mathbf{o}(t)$  is the observation at time  $t$  and  $L_{jm}(t)$  is the occupation probability for the mixture  $m$  of state  $j$  at time  $t$ . If the occupation likelihood of a Gaussian component ( $N_{jm}$ ) is small, then the mean MAP estimate remains close to the mean of initial acoustic model. With MAP adaptation every single component in the system is updates with MAP estimate, based on the prior mean, the weighting and the adaptation data. One drawback to MAP adaptation is that it requires more adaptation data to be effective when compared with other adaptation methods.

### B. MLLR

MLLR [1], [3], [4] computes a set of transformations that will reduce the mismatch between initial model and adaptation data. More specifically MLLR is a model adaptation technique that estimates a set of linear transformations for the mean and variance parameters of acoustic model. Effect of these transformations is to shift the means and alter variances in the initial system. Transformation matrix used to give new estimate of the adapted mean is given by

$$\hat{\mu} = \mathbf{W}\xi \quad (4)$$

where  $\mathbf{W}$  is the transformation matrix and  $\xi$  is the extended mean vector,

$$\xi = [w \mu_1 \mu_2 \dots \mu_n]^T$$

where  $w$  represents a bias offset.  $\mathbf{W}$  can be decomposed into

$$\mathbf{W} = [\mathbf{b}\mathbf{A}] \quad (5)$$



where  $\mathbf{A}$  represents an transformation matrix and  $\mathbf{b}$  represents a bias vector. Transformation of the covariance matrix is in the form of

$$\hat{\Sigma} = \mathbf{H}\Sigma\mathbf{H}^T \quad (6)$$

where  $\mathbf{H}$  is the covariance transformation matrix. CMLLR (Constrained MLLR) originates from MLLR but it is a feature adaptation technique that estimates a set of linear transformations for the features. The effect of this transformation is to shift the feature vector in the initial system so that each state in HMM is more likely to generate adaptation data. The transformation matrix used to give a new estimate of the adapted mean is given by

$$\hat{\delta} = \mathbf{W}\boldsymbol{\varsigma} \quad (7)$$

where  $\mathbf{W}$  is the transformation matrix and  $\boldsymbol{\varsigma}$  is the observation vector that also contain bias offset  $w$  defined as

$$\boldsymbol{\varsigma} = [w \ o_1 \ o_2 \ \dots \ o_n]^T$$

### C. STC

STC [5] models the covariance of the  $m$ -th Gaussian component as

$$\hat{\Sigma}_m = \mathbf{A}^{-1}\Sigma_m\mathbf{A}^{-1T} \quad (8)$$

and mean as

$$\hat{\mu}_m = \mathbf{A}^{-1}\mu_m \quad (9)$$

Single STC matrix is in fact linear transformation applied in feature space that can be estimated using row-by-row iteration scheme. Typical training scheme might start with a single Gaussian system and single  $\mathbf{A}$  matrix. The system is then iteratively refined by re-estimating the component parameters, updating  $\mathbf{A}$  matrix and mixing up until the required number of Gaussians per state are achieved. Gaussian components can be clustered and assigned one  $\mathbf{A}$  matrix per cluster. Simultaneous optimization of the full set of STC parameters is equivalent to maximizing equation

$$\Psi_{STC} = \sum_{t,m} \gamma_m(t) \log \left( \frac{|A|^2}{|diag(AW^m A^T)|} \right) \quad (10)$$

where  $\gamma_m$  is the probability of model occupying mixture component  $m$  and  $W^m$  is the covariance.

### D. HLDA

If each Gaussian component is regarded as a class, then  $W^m$  is the within-class covariance and (10) is the maximum-likelihood solution to generalized form of HLDA [5], in which class covariances are not constrained to be equal. The matrix  $\mathbf{A}$  can be then regarded as feature space transform that discriminates Gaussian components as

$$\hat{\mathbf{y}} = \mathbf{A}\mathbf{y} = \begin{pmatrix} A_{[p]}\mathbf{y} \\ A_{[d-p]}\mathbf{y} \end{pmatrix} = \begin{pmatrix} \hat{\mathbf{y}}_{[p]} \\ \hat{\mathbf{y}}_{[d-p]} \end{pmatrix} \quad (11)$$

where  $d$ -dimensional feature space is divided into  $p$  useful and  $d - p$  nuisance dimensions. The matrix  $\mathbf{A}$  projects  $\mathbf{y}$  into a  $p$ -dimensional subspace that is modelled by diagonal Gaussian mixture components of the acoustic models. The

$d - p$  dimensions are modelled by a global nondiscriminating Gaussian. The optimal value for  $A_{[p]}$  can be estimated by the same row-by-row iteration used in STC case. Equation (10) can be expressed as

$$\Psi_{STC} = \sum_{t,m} \gamma_m(t) \log \cdot \left( \frac{|A|^2}{|diag(A_{[p]}W^m A_{[p]}^T)| |diag(A_{[d-p]}T A_{[d-p]}^T)|} \right) \quad (12)$$

where  $T$  is a global covariance of the training data.

## II. ADAPTATION PROCEDURE

The preparation of IRKR logs for training acoustic models was described in [6]. As in regular recognizer development the first stage in adaptation involves data preparation. Assuming that the file `list.scp` is the list of the source and output files for the adaptation, then they can be coded using the HCopy:

```
HCopy -C genfea.cfg -S list.scp
```

where `genfea.cfg` specifies conversion parameters (frame period, output format, output feature kind and others) [7]. As the output feature kind were chosen MFCCs with delta and delta-delta coefficients. The final stage of data preparation involves generating phone transcriptions of the adaptation data for use in adapting models. The phone level transcriptions were obtained by using HVite to perform forced alignment:

```
HVite -C config -H initiate.model
-i adaptPhones.mlf -I adaptWords.mlf
-S adapt.scp dictionary.dic phones.list
```

where is assumed availability of word level transcriptions (`adaptWords.mlf`), list of coded adaptation data (`adapt.scp`), phonetic dictionary and list of phonemes or triphones.

The supervised adaptation is performed offline by HERest to estimate transformed model, that reduces mismatch between initiate model and the adaptation data. Regression class tree can be used to specify number of transformations to be generated, or the number may be pre-determined using a set of baseclasses. The tool HHed was used to build regression class tree and store it along with set of baseclasses:

```
HHed -H initiate.model -M classes
regtree.hed phones.list
```

HHed creates regression class tree using the initiate model and stores the regression class tree and baseclasses in the `classes` directory. The `regtree.hed` is edit script that contains state occupation statistics file generated by last application of HERest that created the `initiate.model` and commands for building regression tree.

The last step of supervised offline adaptation process is using HERest to estimate transformations using regression class tree. After estimating the transforms, HERest can output either adapted model set or the transformations themselves in transform file or as a set of distinct transformations. Generally it can be invoked via the command

```
HERest [options] -H initiate.model
-u [flags] -J classes -M adapted.model
```

where `[options]` contains transformation specific options, `-u [flags]` specify parameters to be updated, `-J` specifies

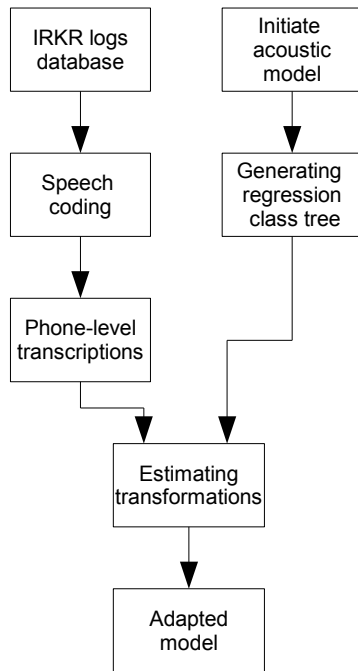


Fig. 1. Adaptation procedure

directory that contains regression class tree and baseclasses and the parameter `-M` specifies output adapted acoustic model. To evaluate the performance of the adapted model is used tool `HVite` and results of the adapted model are observed using `HResults`. There were adapted three sets of acoustic models. First set of acoustic models contains tied-state triphones that were trained on Speechdat database (high-quality speech), other set contains tied-state triphones trained on Mobildat database (contains noises events) and third set contains phoneme models trained also on Mobildat database. These three sets were adapted to IRKR logs database. The adaptation procedure is illustrated in figure (1).

### III. EXPERIMENTAL RESULTS

Adapted acoustic models were tested on Mobildat database on own names. Acoustic models are HMM-based left-to-right 3 state models based on 32 Gaussian mixtures per state. The recognition accuracy is scored by WER (Word Error Rate).

TABLE I  
WER VALUES FOR SET OF TIED-STATE TRIPHONES MODEL TRAINED ON SPEECHDAT DATABASE

tied_32_sd			
base	MAP	STC	HLDA
92.36%	92.36%	90.02%	89.93%

TABLE II  
WER VALUES FOR SET OF TIED-STATE TRIPHONES MODEL TRAINED ON MOBILDAT DATABASE

tied_32_2_md			
base	MAP	STC	HLDA
88.62%	88.62%	89.17%	89.17%

where "base" are WER results for initiate model, "MAP", "STC", "HLDA" are results for model adapted using MAP, STC or HLDA adaptation technique.

TABLE III  
WER VALUES FOR SET OF PHONEME MODELS TRAINED ON MOBILDAT DATABASE

mono_32_md			
base	MAP	STC	HLDA
81.00%	81.00%	63.78%	63.78%

### IV. CONCLUSION

The results of WER achieved with adapted models did not have significant impact on the recognition accuracy of the system. This can be caused by wide environment variability in adaptation database. Other caused is unavailability of enough adaptation data that was acknowledged during the adaptation procedure by the warnings of insufficiency of data, especially in adaptation of triphone models.

### ACKNOWLEDGMENT

The work presented in this paper was supported by Ministry of Education of Slovak Republic under research projects AV 4/0006/07 and Slovak Research and Development Agency under research project APVV-0369-07.

### REFERENCES

- [1] M. Papco, "Robustné metódy automatického rozpoznávania plynulej reči," Ph.D. dissertation, FEI TU Košice, 2008, (in Slovak).
- [2] T. Dat, K. Takeda, and F. Itakura, "The map and cumulative distribution function equalization methods for the speech spectral estimation with application in noise suppression filtering," in *ITRW on Non-Linear Speech Processing (NOLISP 05)*, Barcelona, Spain, 2005, pp. 259–268.
- [3] R. Gomez, T. Toda, H. Saruwatari, and K. Shikano, "Rapid unsupervised speaker adaptation using single utterance based on mllr and speaker selection," in *8th Annual Conference of the International Speech Communication Association (INTERSPEECH 2007)*, Antwerp, Belgium, 2007, pp. 262–265.
- [4] S. Doh, "Enhancements to transformation-based speaker adaptation: Principal component and inter-class maximum likelihood linear regression," Ph.D. dissertation, CMU Pittsburgh, July 2000.
- [5] J. Benesty, M. Sondhi, and Y. Huang, *Handbook of Speech Processing*. Springer-Verlag, 2008.
- [6] M. Papco and J. Juhr, "Acoustic models trained on speechdat database extended to utterances recorded from users interaction with irkr system," in *RTT 2008 : Research in Telecommunication Technology : 9th international conference*, Vyhne, September 10-12 2008.
- [7] S. Young, G. Evermann, M. Gales, T. Hain, D. Kershaw, X. Liu, G. Moore, J. Odell, D. Ollason, D. Povey, V. Valtchev, and P. Woodland, *The HTK Book (for HTK Version 3.4)*, revised for htk version 3.4 ed., December 2006, first published December 1995.

# Bi-parametric model of renewal theory

Peter POÓR

Dept. of Management and Economics, SjF TU of Košice, Slovak Republic

peter.poor@tuke.sk

**Abstract**— During the performance of the company, manufacturing machines and devices need maintenance. With this maintenance, costs are connected. Maintenance guarantees initial attributes of manufacturing machines and devices (conservation), removal of damaged parts or the whole objects (repair). There are two kinds of maintenance: repressive (supplementary) maintenance, where caused damages and failures are being removed, or preventive maintenance, which is executed regardless of the state of the machines, in regular intervals.

Failures, processes of depreciation or time of total abolishment of machines are accidentals. We can not describe properly the whole process of depreciation. When viewing the processes of activity and renewal from shorter time horizon, we apply value point of view. During longer time horizon we need to take time factor into account. Quantitative aspects of renewal process are examined by renewal theory. This contribution presents bi-parametric model, where several criteria are taken into account.

**Keywords**— bi-parametric model, devices, manufacturing, renewal theory

## I. INTRODUCTION INTO THE TOPIC

The renewal theory evolved from actuarial mathematics. Theories of probability were used, for the first time by US actuarial mathematician Alfred James Lotka[1]. The theory has not specified itself very well. We can define renewal as “*a phenomenon consisting in restoration of abilities of object to fulfill desired functions according to technical conditions after failure*”. Later, renewal was understood as a „summary of purposeful activities leading to maintain, or restore object’s ability to work.” The most recent definition describes renewal as a “*phenomenon, when object after breakdown state regains the ability to complete the desired function.*” [2]

Generally, we can say renewal is a process of systematic change of objects (machines, objects, machine parts ...). They need to be replaced with new ones as a consequence of some event (depreciation, sudden failure...). Renewal models examine patterns between objects states, their value, deterioration and maintaining in productive state. If the number of objects is constant, renewal is called easy. If the number of object is during the renewal increasing, we speak about expanded renewal.

## II. BASIC TASK TYPES

- Renewal of a manufacturing machines, whose utility is decreasing. Reasons for replacement of the manufacturing machines are:
  - low performance
  - moral obsolescence
  - high costs for maintenance and repairs
  - combination of above
- Renewal of a manufacturing machines, which suddenly fail and stop fulfill their function
  - system is repaired, only if some component fails
  - control in regular terms, worn-out components are changed
  - all components of this type are changed regularly [3]

Also, renewal models can be divided according to:

- Non-continuous or linear course of renewal – models with discrete or linear time
- Occurrence of accidental values – stochastic or deterministic
- Models of repaired or unrepaired objects
- Renewal costs
- Technically homogenous set of objects or non-technically homogenous set of objects renewal models
- Initial vector structure

The optimal renewal strategy is determined upon operating expenses and maintenance costs comparison of old device with acquisition costs, or different renewal ways costs comparison.

Basically, we can distinguish two types of renewal models:

1. deterministic models, which examine renewal of worn-out devices – products are scrap because of influence of deteriorating. Performance is dropping, but costs are more and more high (cost for maintenance, repairs, energy...). There diseases can be repaired, but maintenance costs lead to value fall. This is the case for example of cars, televisions ...
2. stochastic models, where objects are replaced because of failure – abrasion of the object is presented by probability of object failure. This failure can cause large expenses.

### III. OPTIMALISATION OF RENEWAL TIME

We can distinguish several factors at every manufacturing device:

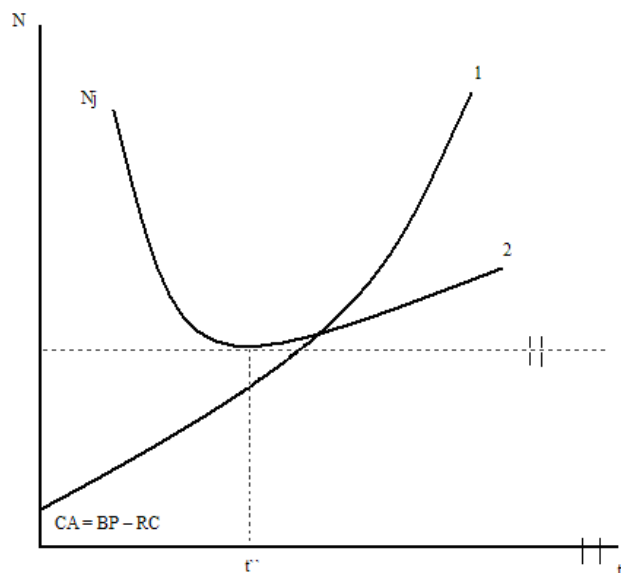
- initial state, when the device has its initial attributes
- period of use, when the state of device is changing depending on time, pieces produced ...
- period of use, when worn-out parts are repaired or replaced
- period of growing expenses
- final period when the device is put aside

We use Selivan's model as a base for manufacturing machines and devices physical depreciation analysis. We start from these assumptions:

- cost of acquisition of the device, where cost of acquisition  $CA = BP - RC$  (buying price - residential cost)
- concrete values for run and repairs
- abilities to replace the device with newer one

Line 1 represents total costs:  
 $N = A + P_1(t) + P_2(t) = A + Bt + Ct^\delta$  (1)

from which we get line 2 (unit costs):  
 $n_j = \frac{N}{t} = \frac{A}{t} + B + Ct^{\delta-1}$  (2)



Pic. 1 Selivan's model of renewal theory

The optimal renewal time is achieved, when unit costs are minimal (graphically tangent to  $N_j$ ). Mathematically it is the first derivation of the equation equal to zero and we get:

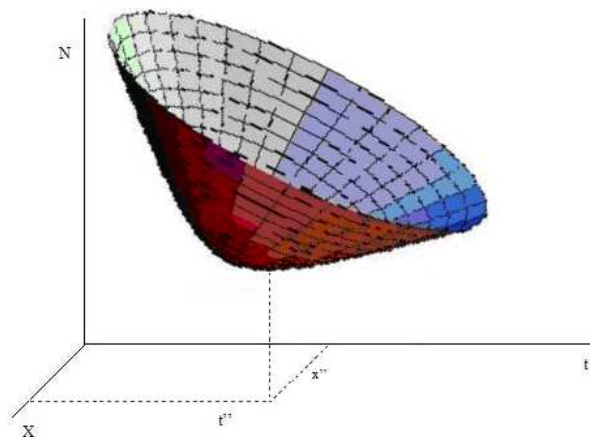
$$t = t_{opt} = \sqrt[\delta]{\frac{A}{(\delta-1) \cdot C}} \quad (3)$$

the optimal renewal time. We see, that this method is usable only when  $\delta > 1$ , when the function  $n_j$  has a local minimum (costs of repairing progressively rise with operation

time).

The insufficiency of this model is that the renewal time depends on one variable. In reality, renewal depends not only from time  $t$ , but also for example from the number of parts produced by the machine, tons of material processed, kilometers passed ... That why it is needed to add another parameters, what makes our model more truthful.

That's why we introduce a bi-parameter model, which has a shape of asymmetric hyperboloid. The minimum of unit costs is assigned at  $t^*$  and  $x^*$ . Now we can see (at constant costs function) the capacity utilization of manufacturing machines and devices and then change the optimum and renewal time.



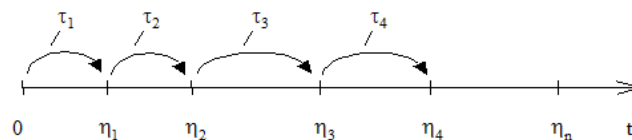
Pic. 2 Bi-parametric model of renewal theory

### IV. RENEWAL THEORY[4]

Failed component are during the renewal process repaired. The renewal process can be defined as a sequence of state without failure and states of failure of random process. We distinguish two types of processes.

*Process with direct renewal, where time of renewal is minimal, or approaching zero.*

Let  $\eta$  be the failure,  $\tau_i$  time of working without failure. Graphically, we can sketch this situation:



$$\eta_n = \sum_{i=1}^n \tau_i \quad (4)$$

Running time of the machine for  $n-1$  recoveries equals to  $n$ -th failure. If the components after renewal are identical, it holds distribution function, created by multi convolution

$$\text{integral: } F_n(t) = P(\eta_n \leq t) = \int_0^t F_{n-1}(t-\tau) dF(\tau) \quad (5)$$

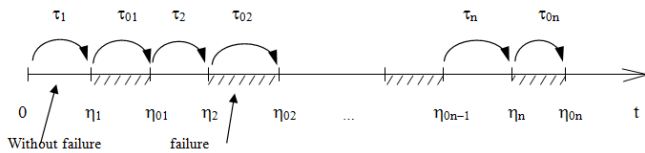
where:  $F_1(t) = F(t)$  is distribution function of failure for single component. The derivation of distribution function is

$$\text{density: } f_n(t) = \int_0^t f_{n-1}(t-\tau)f(\tau)d\tau \quad (6)$$

where:  $f_1(t) = f(t)$  is probability density of failure for single component.

*Process with finite renewal time (renewal time is comparable to time of working without failure)*

Process run:



Time of n-th failure begin:

$$\eta_n = \tau_1 + \tau_{01} + \tau_2 + \tau_{02} + \dots + \tau_n \quad (7)$$

Time of n-th failure end:

$$\eta_{0n} = \tau_1 + \tau_{01} + \tau_2 + \tau_{02} + \dots + \tau_n + \tau_{0n} \quad (8)$$

$$\text{Total time working without failure: } T_p = \tau_1 + \tau_2 + \dots + \tau_n \quad (9)$$

$$\text{Total renewal time: } T_0 = \tau_{01} + \tau_{02} + \dots + \tau_{0n} \quad (10)$$

### V. EXAMPLE

The renewal theory and optimal renewal time determination will be illustrated on this example. A car ŠKODA is owned by a businessman. Cost of acquisition is 250 000 Sk and residential cost is 32 800 Sk.

The optimal renewal time is represented for two cases (to achieve bi-parametric more accurate model):

- a) interdependence of months of use (sequenced into quarters) and cumulative costs
- b) interdependence of kilometers ridden and cumulative costs

*a) interdependence of months of use (sequenced into quarters) and cumulative costs*

Input data are in the table below:

Cumulative costs $N_i$ (€)	Age of car $t_i$ (in quartals)	Distance travelled $x_i$ (in km)
8375,56	1.	895
9344,19	2.	2685
9460.20	3.	4163
9780.56	4.	7540
....	....	....
60307,14	28.	270762
62650,13	29.	238716
68496,98	30.	297656

73248,12	31.	315206
----------	-----	--------

Using the relation (2) with the known data, then simulating with all the data from the table we got these results:  $A = 8740,76$ ;  $B = 171,56$ ;  $C = 0,84$   $\delta = 3,263$  so the cost equation looks like  $N = 8740,76 + 171,56t + 0,84t^{3,263}$  We also calculated  $t_{opt} = 13.276$

*b) interdependence of months of use (sequenced into quarters) and cumulative costs*

Cumulative costs $N_i$ (€)	Distance traveled $x_i$ (in 1 000 km)
9 723	1. (1000 km)
8 886	2. (2000 km)
12 331	3. (3 000 km)
12 835	4. (4 000 km)
...	...
61 800	28. ( 28 000 km)
64 000	29. (29 000 km)
69 030	30. (30 000 km)
73 120	31. (31 000 km)

Using the relation (2) with the known data, then simulating with all the data from the table we got these results:  $A = 9425,71$ ,  $B = 0,07$ ,  $C = 0,00011$   $\delta = 0,0003$ , so the cost equation looks like  $N = 9425,71 + 0,07t + 0,00011t^{0,0003}$ . We also calculated  $t_{opt} = 148919$  km.

### VI. CONCLUSION

From the results of the first case the optimal renewal time is 13,279 (beginning of 13th quarter. This value seems to be good, also according to [5] say, that new car has to be changed every 3-4 years. The last measurement in the example (31 th quarter) represents 8th year of use of the car. The renewal time has been reached two times; it is needed to be changed. In the second case is 148 919 kilometers. According to this, the car should be changed in 18th quarter.

Both values differ, because in first case we look onto moral and material obsolescence of the car, which can be affected only partly (we do not regard whether the device is in service or not). In the second case (interdependence of months of use and cumulative costs) we can affect the moral and material obsolescence (it is realized only when the car is in service).

The best solution is to use the value from b) as the last possible value for renewal and watch the moral and material obsolescence of the automobile (change it according to

optimal renewal time calculated from interdependence of cumulative costs and age of the device).

In the real life, this decision is made by the owner of the device.

*This contribution was created within AV 4/005/07: Use of logistic networks at restructuring of company processes in SME's grant solution.*

#### REFERENCES

- [1] UNČOVSKÝ, L.: Stochastické modely operačnej analýzy. 1. vydanie, Bratislava: ALFA, 1980
- [2] RYBÁŘOVÁ, M.: Diploma thesis – Aplikácia postupov manažérskej logistiky, thesis leader: ŠEBO Dušan
- [3] <https://quercus.kin.tul.cz>, 13.3.2009
- [4] <http://labe.felk.cvut.cz>, 13.3.2009
- [5] <http://www.inhed.logistika.cz>, 14.3.2009

# Modelling of systems with hybrid dynamics

Luboš POPOVIČ

Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

lubos.popovic@tuke.sk

**Abstract**—This paper describes theoretical basis for modelling of hybrid systems, which involve continuous and discrete dynamics and also describes new simulation tools for modelling of systems with hybrid dynamics. Also describes mathematical models of hybrid systems and simulation results of modelling of the specific example of hybrid system.

**Keywords**—Dynamical system, Dynamics, Hybrid system, HYSDEL, Modelling, MPT Toolbox.

## I. INTRODUCTION

Hybrid systems are currently the most discussed problems of control theory. Thus, hybrid systems are systems that involve continuous and discrete variables. Existence of both types of variables, continuous and discrete gives the system hybrid character. Evolution of hybrid systems can be described by using equation, which contains a mixture of logic, discrete and continuous variables. The continuous dynamics of such systems may be continuous-time, discrete-time, or mixed, but is generally given by differential equations. The discrete-variable dynamics of hybrid systems is generally governed by a digital automaton, or input-output transition system with a countable number of states.

## II. MATHEMATICAL DESCRIPTION OF HYBRID SYSTEMS

### A. General Hybrid Dynamical System

Mathematical models reproduce the behavior of physical phenomena. By considering the process at different levels of detail, different models of the same process are usually available. Models should not be too simple, otherwise they do not capture enough details of the process, but also not too complicated in order to formulate and efficiently solve interesting analysis.

In the last years, several computer scientists and control theorists have investigated models describing the interaction between continuous dynamics described by differential or difference equations, and logical components described by finite state machines, if-then-else rules, propositional and temporal logic [1].

Briefly, a hybrid dynamical system is an indexed collection of dynamical systems along with some maps for “jumping” among them. This jumps occurs whenever the state satisfies certain conditions, given by its membership in a specified subset of the state space.

More formally, a general hybrid dynamical system is a system [2]:

$$H = [Q, \Sigma, A, G], \quad (1)$$

with its constituent parts defined as follows:

- $Q$  - the set of index states (discrete states);
- $\Sigma = \{\Sigma_q\}_{q \in Q}$  - collection of constituent dynamical systems, where each  $\Sigma_q = [X_q, f_q]$  is a dynamical system as above with continuous state spaces  $X_q$  and continuous dynamics  $f_q$ ;
- $A = \{A_q\}_{q \in Q}$ ,  $A_q \subset X_q$  for  $\forall q \in Q$  - collection of autonomous jump sets;
- $G = \{G_q\}_{q \in Q}$ , where  $G_q : A_q \rightarrow U_{q \in Q} X_q \times \{q\}$  - collection of jump transition maps, which represent the discrete dynamics of the hybrid system.

As it seen from the description of hybrid systems, which are using two different approaches, let's highlight some advantages:

1. Continuous level:
  - Continuous models are developed and used for long time and therefore are more reliable;
  - For many continuous systems has been proposed several algorithms that can achieve desired behavior;
  - Analog link of actuators and sensors.
2. Discrete level:
  - Easy solving of complexity of the systems using discretization;
  - The simplicity resulting from use of discrete algorithms in computer systems, since they run all computation processes in discrete time.

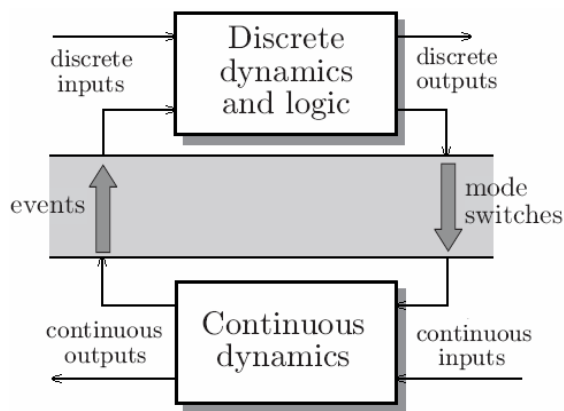


Fig. 1. Block diagram of general hybrid dynamical systems

### B. Discrete Hybrid Automata

Discrete hybrid automata (DHA) are the interconnection of a finite state machine and switched linear dynamic system through a mode selector and an event generator (Fig. 2) [3].

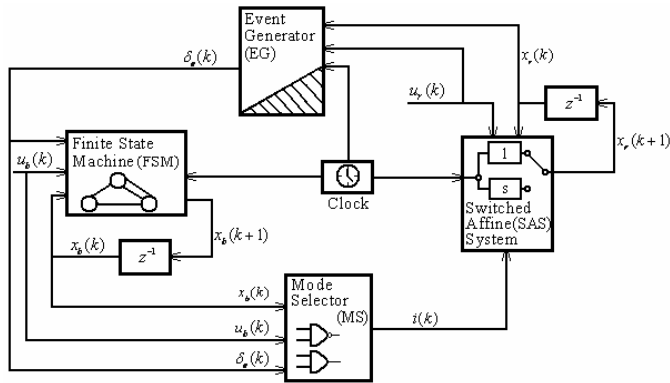


Fig. 2. Block diagram of Discrete hybrid automata

#### Switched Affine System

A switched affine system (SAS) is a collection of linear affine systems:

$$x_r(k+1) = A_{i(k)}x_r(k) + B_{i(k)}u_r(k) + f_{i(k)}, \quad (2)$$

$$y_r(k) = C_{i(k)}x_r(k) + D_{i(k)}u_r(k) + g_{i(k)}, \quad (3)$$

where  $k \in \mathbb{Z}^+$ , is the time indicator,  $x_r \in X_r \subseteq \mathbb{R}^n$  is the continuous state vector,  $u_r \in U_r \subseteq \mathbb{R}^m$  is the exogenous continuous input vector,  $y_r \in Y_r \subseteq \mathbb{R}^p$  is the continuous output vector,  $\{A_i, B_i, f_i, C_i, D_i, g_i\}_{i \in I}$  is a collection of matrices of suitable dimensions, and the mode  $i(k) \in I$  is an input signal that chooses the linear state update dynamics. We denote by  $\#I = s$  the number of elements in  $I$ .

A switched affine system of the form (2) and (3) preserves the value of the state when a switch occurs, however it is possible to implement reset maps on switched affine system, as we will show later in this section. A switched affine system can be rewritten as the combination of linear terms and *if-then-else*. The state-update equation (2) is equivalent to:

$$z_1(k) = \begin{cases} A_1x_r(k) + B_1u_r(k) + f_1, & \text{if } (i(k) = 1), \\ 0, & \text{otherwise,} \end{cases} \quad (4)$$

$\vdots$

$$z_s(k) = \begin{cases} A_sx_r(k) + B_su_r(k) + f_s, & \text{if } (i(k) = s), \\ 0, & \text{otherwise,} \end{cases} \quad (5)$$

$$x_r(k+1) = \sum_{i=1}^s z_i(k), \quad (6)$$

where  $z_i(k) \in \mathbb{R}^n$ ,  $i = 1, \dots, s$ . Equation (3) admits a similar transformation.

#### Event Generator

An event generator (EG) is a mathematical object that generates a logic signal according to the satisfaction of a linear affine constraint:

$$\delta_e(k) = f_H(x_r(k), u_r(k), k), \quad (7)$$

where  $f_H : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{Z}_{\geq 0} \rightarrow D \subseteq \{0, 1\}^n$  is a vector of descriptive functions of a linear hyperplane, and  $\mathbb{Z}_{\geq 0} \triangleq \{0, 1, \dots\}$  is the set of nonnegative integers. In particular, *time events* are modeled as:  $[\delta_e^i(k) = 1] \leftrightarrow [kT_s \geq t_0]$ , where  $T_s$  is the sampling time, and *threshold events* are modeled as:  $[\delta_e^i(k) = 1] \leftrightarrow [a^T x_r(k) + b^T u_r(k) \leq c]$ , where the superscript  $i$  denotes the  $i$ -th component of a vector.

#### Finite State Machine

A finite state machine (FSM) or automaton is a discrete dynamic process that evolves according to a logic state-update function:

$$x_b(k+1) = f_b(x_b(k), u_b(k), \delta_e(k)), \quad (8)$$

where  $x_b \in X_b \subseteq \{0, 1\}^n$  is the Boolean state,  $u_b \in U_b \subseteq \{0, 1\}^m$  is the exogenous Boolean input,  $\delta_e(k)$  is the endogenous input coming from the event generator (EG), and  $f_b : X_b \times U_b \times D \rightarrow X_b$  is a deterministic logic function. A finite state machine can be conveniently represented using an oriented graph and we will only refer to synchronous finite state machine, where the transitions may happen only at sampling times. The adjective synchronous will be omitted for brevity.

A finite state machine may also have an associated Boolean output:

$$y_b(k) = g_b(x_b(k), u_b(k), \delta_e(k)), \quad (9)$$

where  $y_b \in Y_b \subseteq \{0, 1\}^p$ . Figure (Fig. 3) shows a finite state machine where  $u_b = [u_{b1}, u_{b2}]^T$  is the input vector, and  $\delta = [\delta_1, \dots, \delta_4]^T$  is a vector of signals coming from the event generator.

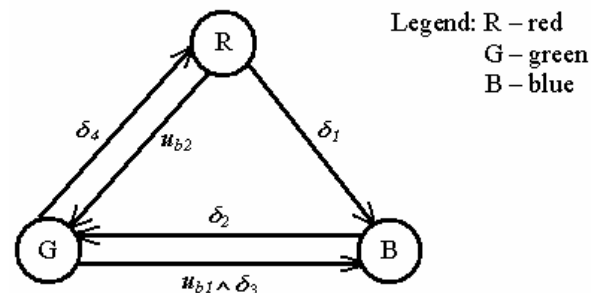


Fig. 3. Example of finite state machine

The state transition function is:

$$x_b(k+1) = \begin{cases} R & \text{if } ((x_b(k) = G) \wedge \delta_4) \vee ((x_b(k) = R) \wedge \neg u_{b2} \wedge \neg \delta_1), \\ G & \text{if } ((x_b(k) = R) \wedge u_{b2}) \vee ((x_b(k) = B) \wedge \delta_2) \vee \\ & ((x_b(k) = G) \wedge \neg \delta_4 \wedge \neg (u_{b1} \wedge \delta_3)), \\ B & \text{if } ((x_b(k) = R) \wedge \delta_1) \vee ((x_b(k) = G) \wedge (u_{b1} \wedge \delta_3)) \vee \\ & ((x_b(k) = B) \wedge \neg \delta_2). \end{cases} \quad (10)$$



### Mode Selector

The logic state  $x_b(k)$ , the Boolean inputs  $u_b(k)$ , and the events  $\delta_e(k)$  select the dynamic mode  $i(k)$  of the switched affine system through a Boolean function  $f_M : X_b \times U_b \times D \rightarrow I$ , which is therefore called *mode selector*. The output of this function:

$$i(k) = f_M(x_b(k), u_b(k), \delta_e(k)), \quad (11)$$

is called *active mode*. We say that a *mode switch* occurs at step  $k$  if  $i(k) \neq i(k-1)$ . Note that contrarily to continuous time hybrid models, where switches can occur at any time, in our discrete-time setting a mode switch can only occur at sampling instants.

### Reset Maps

In correspondence with a mode switch and  $i(k) = j$ ,  $i(k-1) = h$ ,  $h \neq j$ ,  $h, j \in I$  instead of evolving  $x_r(k+1) = A_j x_r(k) + B_j u_r(k) + f_j$  it is possible to associate a reset of the continuous state vector:

$$x_r(k+1) = \phi_{hj}^-(x_r(k), u_r(k)) = A_{hj}^- x_r(k) + B_{hj}^- u_r(k) + f_{hj}^-, \quad (12)$$

where the function  $\phi_{hj}^-$  is called *reset map*. The reset can be considered as special dynamics that only acts for one sampling step. Let's look on figure (Fig. 4.), that shows how a reset affects the state evolution.

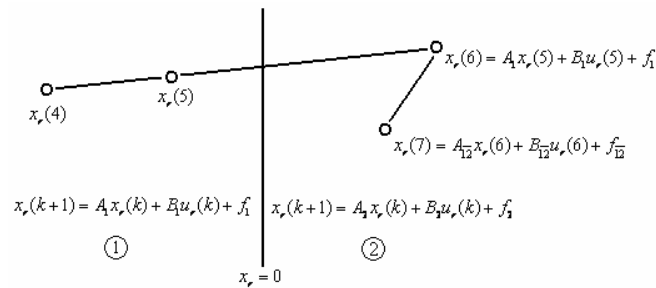


Fig. 4. Reset maps – A posteriori resets

At time  $k = 5$  the system is in mode  $i(5) = 1$ , at  $k = 6$  the state  $x_r(6) = A_1 x_r(5) + B_1 u_r(5) + f_1$  enters the region  $x_r \geq 0$ . This generates an event  $\delta_e(6)$  through the event generator, which in turn causes the mode selector to change the system dynamics to  $i(6) = 2$ . The mode switch  $1 \rightarrow 2$  resets  $x_r(7) = A_{12} x_r(6) + B_{12} u_r(6) + f_{12}$ . If the state  $x_r(7)$  after reset belongs again to the region where the mode 2 is active,  $i(7) = 2$  the successor state is  $x_r(8) = A_2 x_r(7) + B_2 u_r(7) + f_2$ . It might even happen that  $x_r(7)$  belongs to another region, say a region where mode 3 is active,  $i(7) = 3$ . In this case, since  $i(6) \neq i(7)$ , a further reset  $2 \rightarrow 3$  is applied and  $x_r(8) = A_{23} x_r(7) + B_{23} u_r(7) + f_{23}$ .

In some circumstances, it is desirable to predict the mode switch and to anticipate the reset by one sampling step, i.e., to reset the state *before* the guardline is actually crossed.

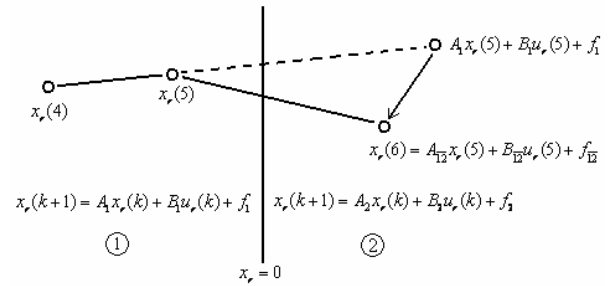


Fig. 5. Reset maps – Predictive resets

Consider figure (Fig. 5), where at time  $k = 5$  the state  $x_r(5)$  and the input  $u_r(5)$  are such that  $A_1 x_r(5) + B_1 u_r(5) + f_1 \geq 0$  which would generate an event  $\delta_e$  at the next time step. As a consequence of the predicted mode switch, the state is reset according to the reset map  $f_{12}^-(x, u)$ , i.e.,  $x(6) = A_{12} x(5) + B_{12} u(5) + f_{12}$ .

### III. SIMULATION EXAMPLE

As an example of system with hybrid dynamics we consider the system of two tanks with liquid as shown in figure (Fig. 6). Dynamics properties of this system are changing over time and are described by system of differential equations. Transition between these dynamics occurs in certain switching time [4].

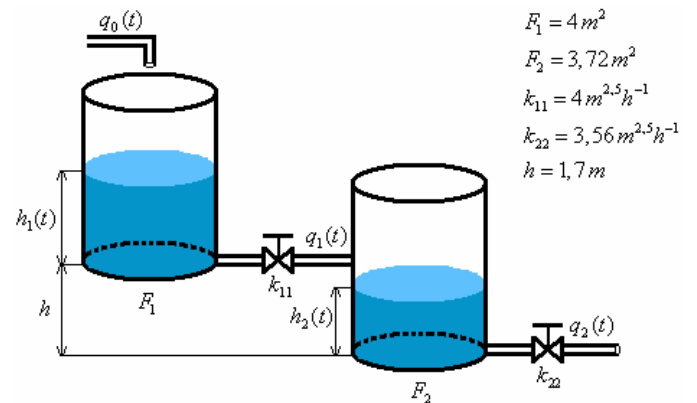


Fig. 6. Hybrid system of two tanks with liquid

Consider figure (Fig. 6), where tanks are in different height. Thus switch between dynamics occurs in moment when level of liquid in second tank exceeds height of bottom of the first tank. In this moment the mathematical model of whole system changes from system without interaction to system with interaction.

To obtain both models we used material balance formulated for weight:

(Sum of mass flow on input) = (Sum of mass flow on output) + (Rate of accumulation of mass in whole system).

Then for system of two tanks without interaction we obtain:

$$q_0(t) = k_{11} \sqrt{h_1(t)} + F_1 \frac{dh_1(t)}{dt} \quad (13)$$

$$k_{11} \sqrt{h_1(t)} = k_{22} \sqrt{h_2(t)} + F_2 \frac{dh_2(t)}{dt} \quad (14)$$

For system of two tanks with interaction we obtain:

$$q_0(t) = \text{sign}(h_1(t) - (h_2(t) - h)) k_{11} \sqrt{|h_1(t) - (h_2(t) - h)|} + F_1 \frac{dh_1(t)}{dt} \quad (15)$$

$$\text{sign}(h_1(t) - (h_2(t) - h)) k_{22} \sqrt{|h_1(t) - (h_2(t) - h)|} = F_2 \frac{dh_2(t)}{dt} + k_{22} \sqrt{h_2(t)} \quad (16)$$

### A. Simulation using MPT Toolbox

Hybrid system shown on figure (Fig. 6) we are able to model and simulate in MPT Toolbox of simulation language Matlab. First we express levels of liquid in both tanks from (13)-(16), then we use a forward difference (17) to obtain discrete form of equations, which we can easily write to MPT Toolbox.

$$\frac{dx(t)}{dt} \approx \frac{x(k+T) - x(k)}{T} \quad (17)$$

Results of simulation on this hybrid system in Matlab/MPT Toolbox are shown on figure (Fig. 7).

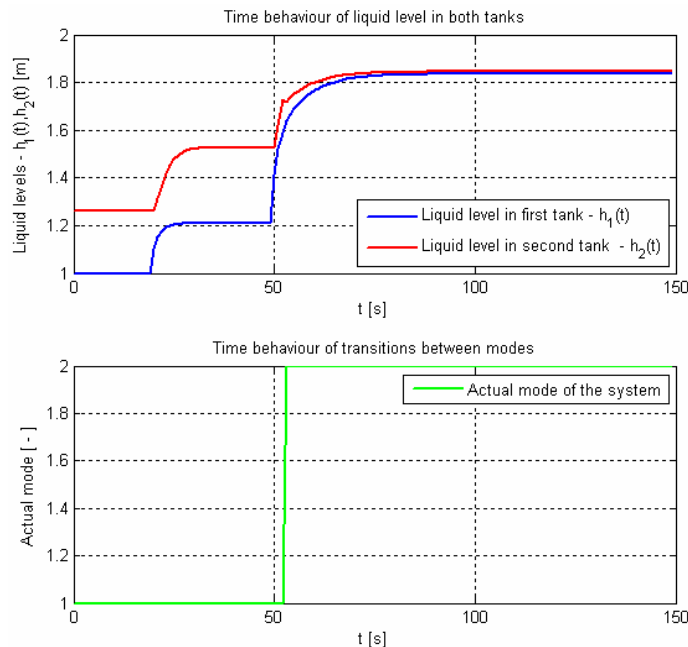


Fig. 7. Time behaviour of liquid level and mode in tanks

### B. Simulation using HYSDEL

In this section for modelling and simulation of hybrid system we used tool HYSDEL (Hybrid System Description Language). First we express levels of liquid in both tanks from (13)-(16) and then these nonlinear differential equations must be linearized about equilibrium, by using Taylor's evolution. After that we use a forward difference (17) to obtain discrete form of linearized equations, which we can easily write to HYSDEL:

1. System without interaction:

$$\Delta h_1(t+T) = \left(1 - \frac{k_1}{F_1}\right) \Delta h_1(t) + \frac{\Delta q_0(t)}{F_1} \quad (18)$$

$$\Delta h_2(t+T) = \frac{k_1}{F_2} \Delta h_1(t) + \left(1 - \frac{k_2}{F_2}\right) \Delta h_2(t) \quad (19)$$

2. System with interaction:

$$\Delta h_1(t+T) = \left(1 - \frac{k_1}{F_1}\right) \Delta h_1(t) + \frac{k_1}{F_1} \Delta h_2(t) + \frac{\Delta q_0(t)}{F_1} \quad (20)$$

$$\Delta h_2(t+T) = \frac{k_1}{F_1} \Delta h_1(t) - \left(\frac{k_1}{F_2} + \frac{k_2}{F_2} - 1\right) \Delta h_2(t) \quad (21)$$

Switching condition in both cases of simulation is  $h_2(t) > h$ .

Results of simulation this hybrid system in HYSDEL/Matlab are shown on figure (Fig. 8).

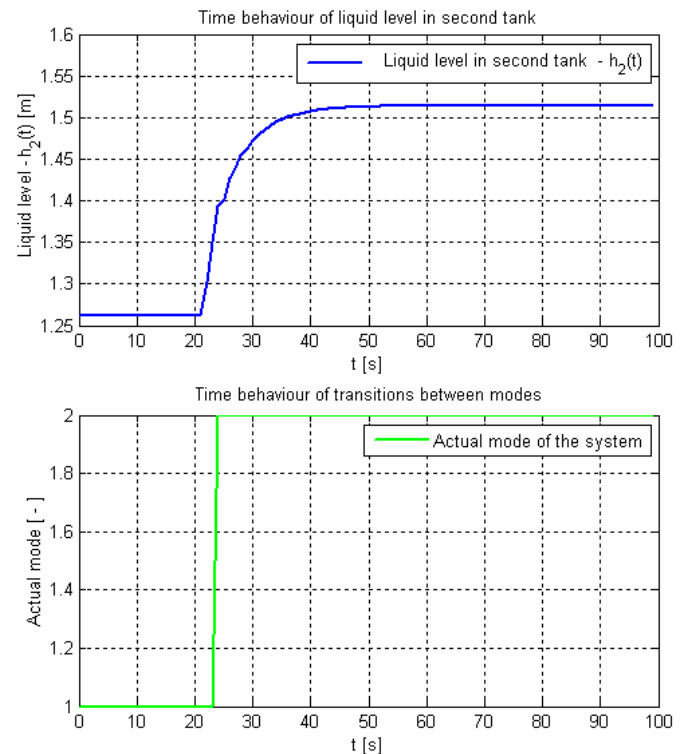


Fig. 8. Time behaviour of liquid level and mode in second tank

### ACKNOWLEDGMENT

The work presented in this paper was supported by Scientific Grant Agency of the Ministry of Education of Slovak Republic and Slovak Academy of Science No. 1/0617/08.

### REFERENCES

- [1] P. J. Antsaklis,.: *A Brief Introduction to Theory and Applications of Hybrid Systems*. In: Proceedings of IEEE, Special Issue on Hybrid Systems: Theory and Applications. vol. 88, no. 7, 2000, pp. 879-887.
- [2] M. S. Branicky: *Studies in Hybrid Systems: Modelling, Analysis, and Control*, dissertation work. Massachusetts, 1995. pp. 198.
- [3] F. D. Torrisi, A. Bemporad: *Discrete-time Hybrid Modelling and Verification*, In. Proc. 40th IEEE Conf. on Decision and Control, Orlando, Florida, 2001, pp. 2899-2904.
- [4] T. Hirmajer – M. Fikar: *Optimal control of system with hybrid dynamics*. In: AT&P Journal. vol. 8, no. 12, 2005, pp. 81-84.
- [5] M. Kvasnica – P. Grieder – M. Baotić – F.J. Christophersen: *Multi-Parametric Toolbox – user manual*. Zürich, Automatic Control Laboratory (Institut für Automatik), 2005. pp. 79
- [6] F. D. Torrisi – A. Bemporad – G. Bertini – P. Hertach – D. Jost – D. Mignone: *HYSDEL 2.0.5 – user manual*, 2002. pp. 35.

# Complete Signal Processing Procedure for Through Wall Target Tracking: Description and Evaluation on Real Radar Data

Jana ROVNÁKOVÁ

Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

jana.rovnakova@tuke.sk

**Abstract**—In this paper, through wall moving target tracking by UWB radar is described as a complex procedure with all required phases of radar signal processing. For particular phases of that process, i.e. for raw radar data pre-processing, background subtraction, detection, trace estimation, localization and tracking itself, the phase significance and applied specific method are outlined. The procedure performance is evaluated on the base of processing results of real data acquired by M-sequence UWB radar.

**Keywords**—Radar signal processing, through wall measurement, target tracking, UWB radar.

## I. INTRODUCTION

Through wall target tracking can be profitable e.g. in an interior inspection for rescue applications (fire, earthquake,...), during security operations (hostage situations, explosion,...) or like surveillance for industrial purposes. It is advantageously realized by ultra-wideband (UWB) radars which operate in a lower GHz-range base-band - approximately up to 5 GHz. Such devices are then characteristic with good penetration of emitted signals through various obstacles, e.g. through most common building materials including reinforced concrete, concrete block, sheet rock, brick, wood, plastic, tile or berglass.

Radar signal processing for moving target tracking, i.e. determining target coordinates as the continuous function of time, is a complex process that includes following phases: raw radar data pre-processing, background subtraction, detection, trace estimation, localization and tracking itself. For realization of every listed task, there are available several classes of different techniques, overviews of which can be found e.g. in [1], [2], [3], [4], [5], [6]. However only rarely they are described all together as individual components of the full process [7].

The intention of this paper is to describe and evaluate the complete procedure of UWB radar signal processing for moving target tracking consisting of specific sequence of methods. Within the frame of it a novel approach for trace estimation is introduced. The performance of proposed procedure has been tested on real data obtained by UWB radar which uses maximum-length-binary-sequence (M-sequence) as stimulation signal [8].

## II. DESCRIPTION OF COMPLETE SIGNAL PROCESSING PROCEDURE

Raw radar data can be interpreted as a set of impulse responses of surrounding through which the emitted signals

were propagated. In the case of M-sequence UWB radar such signals must be at first *pre-processed* by time-zero setting. Time-zero is the exact time instant at which the transmitting antenna starts emitting the first elementary impulse of M-sequence (so-called chip) and depends e.g. on the cable lengths, total group delays of radar device electronic systems, etc., but especially on the chip position at which the M-sequence generator started to generate the first M-sequence. This position is randomly changed after every power supply reconnecting. To find time-zero means rotate all received impulse responses in such a way as their first chips correspond to the spatial position of the transmitting antenna. For finding the number of chips needed for such rotating we utilize cross-talk signal [1].

After phase of pre-processing, a signal to noise ratio of radar data is needed to improve. It is done by *background subtraction* which rejects especially the stationary and correlated clutter such as antenna coupling, impedance mismatch response and ambient static clutter, and allows the response of moving targets to be detected. From variety of background subtraction methods we have chosen exponential averaging [9]. This technique is ranked among popular and often used methods because of its low complexity and no more memory requirements (need only one previous impulse response). Therefore is also suitable for on-line processing.

*Detection* is the next step in the radar data processing which comes after background subtraction. It represents a class of methods that on base of some decision theory determine whether a target is absent or present in examined radar data. Between detectors which are able to provide good and robust results in the case of through wall target detection by UWB radar, a constant false alarm rate (CFAR) detector can be assigned. It is based on Neymann-Person optimum criterion providing the maximum probability of detection for a given false alarm rate. In our processing we have applied CFAR detector that assumed a Gaussian clutter model [10].

Binary data that are output of detector form a noticeable trace of moving targets. It represents time of arrival (TOA) of electromagnetic waves reflected by target for particular time observation instants. As the range resolution of UWB radars is considerably smaller than the physical dimensions of the targets to be detected, the position of targets are usually represented by more TOA estimations. In order to simplify the localization, such distributed targets are replaced by simple targets, i.e. the position of target is in every time observation

instant given by maximum one TOA [11]. This phase of radar signal processing is called *trace estimation* and we have performed it on the base of new algorithm, a description of which is given in the next section.

Target traces obtained from all receiving antennas are in the following phase of the radar data processing used as input parameters for location algorithms. The aim of *localization* is to determine target coordinates in defined coordinate systems whereby target locations estimated in consecutive time instants create target trajectory. As the radar system which was used during our measurements consists from one transmitting and two receiving antennas, only non-iterative direct method of localization could be used. In that case, the target coordinates were simply calculated like intersections of ellipses [12].

The particular locations of the target are estimated with certain random error usually described by its probability distribution function. Taking into account this model of the target position estimation, the target trajectory can be further processed by tracking algorithms. They provide a new estimation of target location based on foregoing positions of the target. Usually, the *tracking* results in the target trajectory error decreasing including trajectory smoothing. From different tracking algorithms a linear Kalman filter has been chosen like the method which encloses proposed procedure of UWB radar signal processing [13].

#### A. Target Trace Estimation

Radar signal processing which includes trace estimation is alternative approach to conventional processing by means of radar imaging [14]. Its advantage is in decreasing of computational complexity because it enables to work with vectors (impulse responses) instead of matrices (radar images). We have gradually developed few algorithms for trace estimation. The presented one is quite simple but enable to process also scenarios with multiple targets. It was programmed for on-line processing of radar data.

The target trace estimation algorithm under consideration can be described in four simplified steps. The aim of the first step is to eliminate the influence of the wrong results of detection, i.e. false alarms and detection results where the target should be detected but it has not been detected. This is done by means of threshold summation of impulse response samples with window length corresponding to assumed size of target.

After this step, the potential targets create continuous sequences of "1", position of which indicate TOA of reflections from different parts of human body. By comparing the computed and measured traces we have found that the best pointed representation of distributed targets is given by their first reflections. Therefore those points represent the simple targets.

During the third step the surrounding of simple targets is examined. Under the term "surrounding", we understand two dimensional window consisting from few samples of forgoing and consequent impulse responses. The aim is to confirm the existence of target on the base target presence in several time observation instants. Dimension of window in the direction of time observation should be chosen carefully because it affects the time delay of whole processing.

The last step consists in smoothing and completing of trace in given target surrounding. The smoothing is done by



Fig. 1. The gymnasium: a) an exterior, b) an interior.

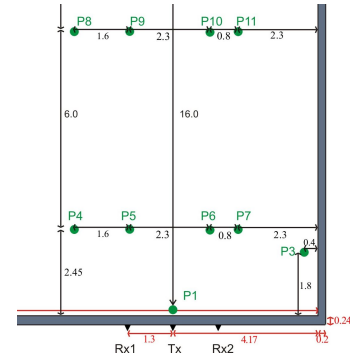


Fig. 2. Scheme of the gymnasium: a person was walking through positions P7-P8-P4-P11-P7.

averaging of circumjacent pointed target positions but only in the case of sufficient number of such positions. For completing of missing TOA in the trace is enough to have confirmation of target in two other time instants. We use linear interpolation for that purpose but only in the left target surrounding which correspond with verified simple targets positions.

### III. EVALUATION OF PROPOSED PROCEDURE ON REAL RADAR DATA

The real data were acquired by M-sequence UWB radar system during measuring campaign in Stockholm Rescue Center. The radar device worked with 4.5 GHz internal clock and included one transmitting ( $Tx$ ) and two receiving antennas ( $Rx_1$ ,  $Rx_2$ ). Measurement speed was 13.44 impulse responses per second. All scenarios with moving targets were measured through walls of different materials and thickness (46 cm brick, 17 + 19 cm concrete, 50 cm concrete, 15 cm light concrete, 50 + 40 cm light concrete, 18 cm massive wood, 8 cm plasterboard, 24 cm tile, 18 cm wood). Due to limited space of this paper, results only for three diverse scenarios with one moving person are described in the next text.

The first chosen measurement can be ranged between simpler scenarios because a person was walking in large empty room (a gymnasium, Fig. 1b). Its trajectory had shape of a ribbon (Fig. 2). The radar system was placed behind 24 cm thick wooden wall covered by tile (Fig. 1a). The output data of every processing phase are depicted in Fig. 7. The reached target track corresponds with true trajectory of moving person (Fig. 7e). Other reason for obtaining such good result was the fact that distance between adjacent antennas was quite high - 130 cm. This positive influenced the accuracy of localization algorithm.

The conditions in the 2<sup>nd</sup> scenario were more complex. The person was walking through narrow corridor upstairs and back (Fig. 3b and Fig. 4). A set up of radar antennas is photographed in Fig. 3a. They were placed behind 17 cm thick concrete wall



Fig. 3. The corridor with stairs: a) a side wall, b) an interior.

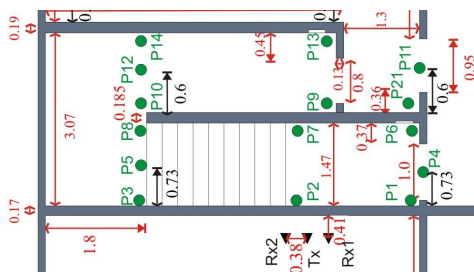


Fig. 4. Scheme of the staircase: a person was walking through positions P4-P5-P4.

with 38 cm distance between adjacent antennas. In processed radar data (Fig. 8b-d) is possible to see multiple reflection caused by the near position of back wall. Despite of this the final track truly copies the person trajectory (Fig. 8e). The lower localization accuracy became evident only in parts corresponding with movement up and down the stairs.

The last scenario is the hardest one. We have measured behind 46 cm thick brick wall in fully furnished room (a classroom, Fig. 5). The trajectory of walking person had shape of a letter "S" (Fig. 6). The distance between antennas was again 38 cm. In the output data from background subtraction, detection and trace estimation are clearly visible reflections like from other target (Fig. 9). They are caused by a blackboard which was placed through all back wall approximately in distance of 10 m from radar system. This effect is known in literature under the term "shadowing" and its impact is shown in Fig. 9e by black stars. The beginning positions of target track were lost due to initial phase of recursively working exponential averaging. Small distortions of obtained track are visible in farther positions what relates with the decreasing magnitude of radar signals. In spite of all mentioned difficulties, we consider the reached target track for satisfying.

#### IV. CONCLUSION

The obtained results illustrate performance of the complete radar signal processing procedure for moving target tracking. The comparison of the true and estimated target tracks confirms correctness of the proposed phases. During evaluation of chosen three scenarios some challenging tasks like multiple reflections and existence of "shadows" was pointed out. Their solutions as well as enhancement of weak signals in the case of multiple target scenarios will be the content of our next research.

#### ACKNOWLEDGMENT

This work was supported by the Slovak Research and Development Agency under the contract No. LPP-0287-06 and by European Commission under the contract COOP-CT-2006-032744.



Fig. 5. The classroom: a) an exterior, b) an interior.

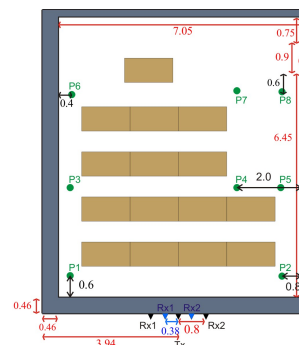


Fig. 6. Scheme of the classroom: a person was walking through positions P1-P2-P5-P3-P6-P7.

#### REFERENCES

- [1] R. Yelf, "Where is true time zero?" in *Proceedings of the Tenth International Conference on Ground Penetrating Radar*, June 2004, pp. 279–282.
- [2] M. Piccardi, "Background subtraction techniques: a review," in *Proceedings of International Conference on Systems, Man and Cybernetics*, The Hague, The Netherlands, October 2004.
- [3] J. D. Taylor, *Ultrawideband radar technology*. CRC Press, 2000.
- [4] D. Stanford and A. E. Raftery, "Principal curve clustering with noise," Dep. of Statistics, University of Washington, Tech. Rep. 317, February 1997.
- [5] K. Yu, H. Saarnisaari, J. P. Montillet, A. Rabbachin, I. Oppermann, and G. Abreu, *Ultra-wideband wireless communications and networks*. John Wiley & Sons, Ltd, February 2006, ch. Localization.
- [6] S. Blackman and R. Popoli, *Design and analysis of modern tracking systems*. Artech House Publishers, 1999, ISBN 1-58053-006-0.
- [7] J. Rovňáková, M. Švecová, D. Kocur, T. T. Nguyen, and J. Sachs, "Signal processing for through wall moving target tracking by m-sequence ubw radar," in *The 18th International Conference Radioelektronika*, Prague, Czech Republic, April 2008.
- [8] J. Sachs, P. Peyerl, and R. Zetik, "Stimulation of uwb-sensors: Pulse or maximum sequence?" in *International Workshop on UWB Systems*, Oulu, Finland, June 2003.
- [9] R. Zetik, S. Crabbe, J. Krajinak, P. Peyerl, J. Sachs, and R. Thoma, "Detection and localization of persons behind obstacles using m-sequence through-the-wall radar," in *Proceedings of SPIE - Sensors, and Command, Control, Communications, and Intelligence Technologies for Homeland Security and Homeland Defense*, vol. 6201, May 2006.
- [10] P. K. Dutta, A. K. Arora, and S. B. Bibyk, "Towards radar-enabled sensor networks," in *Proceedings of the Fifth International Conference on Information Processing in Sensor Networks. Special Track on Platform Tools and Design Methods for Network Embedded Sensors*, Nashville, Tennessee, USA, April 2006, pp. 467 – 474.
- [11] D. Kocur, J. Rovnakova, M. Švecova, J. Sachs, and E. Zaikov, "Midterm report on person detection and localisation," Tech. Rep., 2009, project: Ultra Wideband Radio application for localisation of hidden people and detection of unauthorised objects.
- [12] M. Aftanas, J. Rovnakova, M. Riskova, D. Kocur, , and M. Drutarovsky, "An analysis of 2D target positioning accuracy for M-sequence UWB radar system under ideal conditions," *Proceedings of 17th International Conference Radioelektronika, Brno*, pp. 189–194, Apr. 2007.
- [13] M. S. Grewal and A. P. Andrews, *Kalman filtering: Theory and practice using MATLAB*, 3rd ed. Wiley-IEEE Press, September 2008.
- [14] G. Hanwei, L. Diannong, W. Yan, H. Xiaotao, and D. Zhen, "Moving target imaging using ultra-wide band synthetic aperture radar," in *Proceedings of SPIE: Algorithms for Synthetic Aperture Radar Imagery, Vol. 5095*, 2003, pp. 233 – 241.

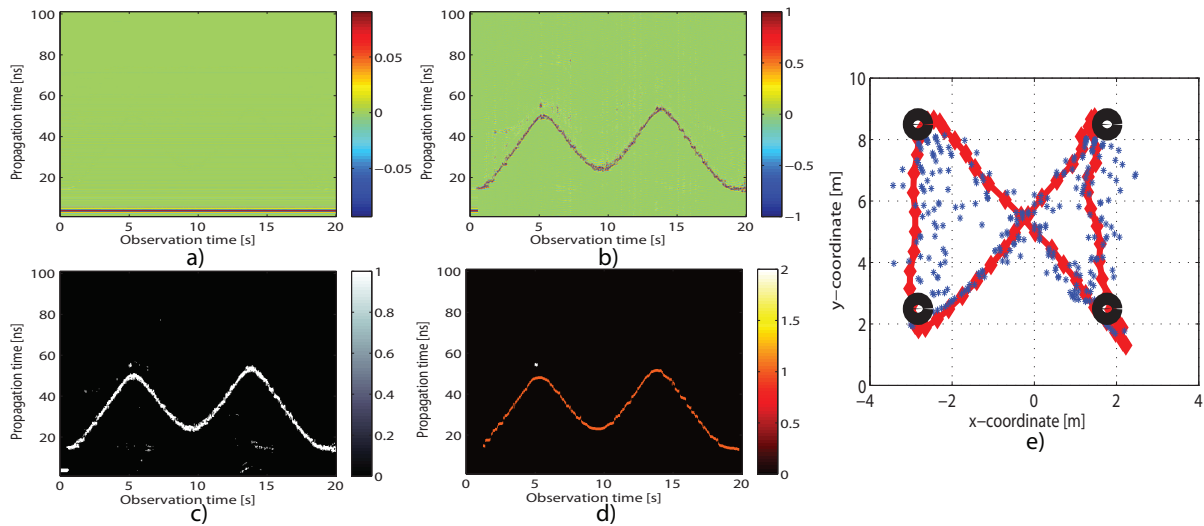


Fig. 7. SCENARIO 1 - the gymnasium covered by tile. The output data after phase of: a) *pre-processing* - signals with original magnitude from  $Rx_1$ , b) *background subtraction* - normalized signals from  $Rx_1$ , c) *detection* - binary data from  $Rx_1$ , d) *trace estimation* - the trace color designates the number of targets, data from  $Rx_1$  e) *localization* (blue stars) and *tracking* (red diamonds) - black circles denote positions through which the target was moving.

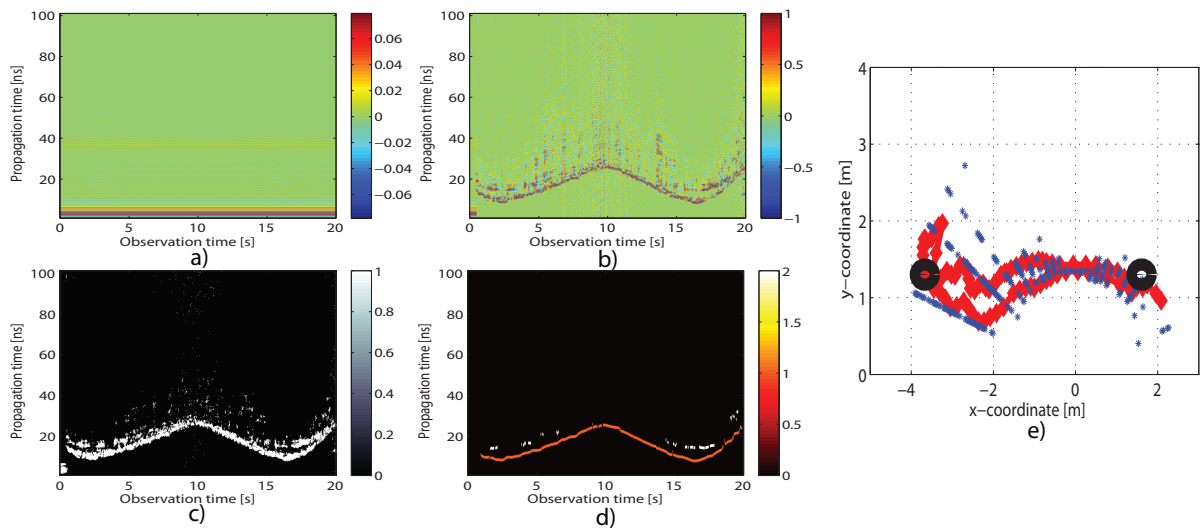


Fig. 8. SCENARIO 2 - the concrete staircase. The output data after phase of: a) *pre-processing* - signals with original magnitude from  $Rx_1$ , b) *background subtraction* - normalized signals from  $Rx_1$ , c) *detection* - binary data from  $Rx_1$ , d) *trace estimation* - the trace color designates the number of targets, data from  $Rx_1$  e) *localization* (blue stars) and *tracking* (red diamonds) - black circles denote positions through which the target was moving.

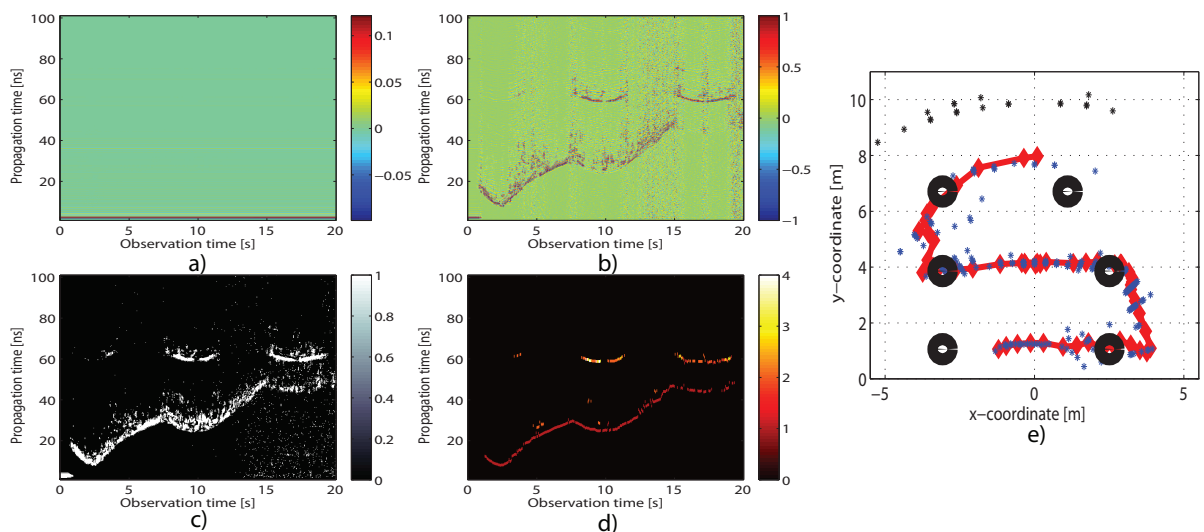


Fig. 9. SCENARIO 3 - the brick classroom. The output data after phase of: a) *pre-processing* - signals with original magnitude from  $Rx_1$ , b) *background subtraction* - normalized signals from  $Rx_1$ , c) *detection* - binary data from  $Rx_1$ , d) *trace estimation* - the trace color designates the number of targets, data from  $Rx_1$  e) *localization* (blue and black stars) and *tracking* (red diamonds) - black circles denote positions through which the target was moving.

# Soft switching PS-PWM DC/DC converter using auxiliary circuit

<sup>1</sup>Vladimír Ruščin, <sup>2</sup>Marcel Bodor

Department of Electrical, Mechatronics and Industrial Engineering, FEI TU of Košice, Slovak Republic

<sup>1</sup>vladimir.ruscin@tuke.sk, <sup>2</sup>bodorm@orangemail.sk

**Abstract**— A novel soft switching PS-PWM DC/DC converter with controlled secondary side rectifier using secondary energy recovery snubber is presented in this paper. Soft switching for all power switches of the converter is achieved for full load range from no-load to short circuit by using controlled rectifier and snubber on the secondary side. Modified phase shift PWM control strategy is used for the converter. The principle of operation is explained and analyzed and experimental results on the laboratory model are presented.

**Keywords**— auxiliary circuit, soft switching, zero voltage zero current switching.

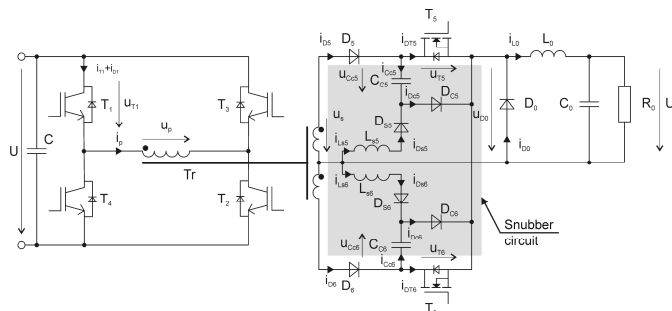


Fig. 1 Power scheme of the DC/DC converter

## I. INTRODUCTION

A controlled inverter is the conventional way to control the output value in the conventional PWM converters. Secondary side of the power transformer is usually fitted by a non-controlled center tapped or bridge rectifier. This topology of DC/DC converters permits all switching devices to operate under zero-voltage switching by using circuit parasitics such as power transformer leakage inductance and devices junction capacitance. Because of phase shifted PWM control, the converter has a disadvantage that circulating current flows through the power transformer and switching devices during freewheeling intervals. This circulating current can be eliminated by application of the reverse bias, or secondary controlled rectifier.

## II. POWER SCHEME OF THE CONVERTER

To improve the properties of the existing converters, the new topology of PWM DC/DC converter was developed. The proposed DC/DC converter shown in Fig. 1 consists of high-frequency full bridge inverter, center tapped power transformer, controlled output rectifier, output filter and novel type of secondary energy recovery snubber. The converter is controlled by modified pulse width modulation (Fig. 2). With this modification of the phase shift PWM control it was achieved that primary transistors turn on under zero current conditions, and turn off with negligible losses. The new snubber circuit eliminates the turn off losses of the secondary transistors because turn off under zero voltage is ensured. The semiconductor switches  $T_5$ ,  $T_6$  in the secondary side are used to reset secondary and simultaneously also primary circulating current

## III. OPERATION PRINCIPLE

The switching diagram and operation waveforms are shown in Fig. 2. The basic operation of the proposed soft switching converter consists of eight operating modes (intervals) within each half cycle.

**Interval ( $t_0$ - $t_1$ ):** At the time  $t_0$  the primary transistors  $T_1$ ,  $T_2$  are turned on. The leakage inductance of the power transformer ensures that the emitter current of these transistors rises with a reduced  $di/dt$ , so the IGBT switches are turned on under zero current conditions. The rectifier switch  $T_6$  is turned on in the same interval and the capacitor  $C_{C6}$  is discharging to load through smoothing choke. At the end of this interval is the capacitor  $C_{C6}$  fully discharged.

**Interval ( $t_1$ - $t_2$ ):** During this period the secondary current rises with a slope that depends on the output filter inductance  $L_0$ .

**Interval ( $t_2$ - $t_3$ ):** At the time  $t_2$  the secondary transistor  $T_5$  is turned off. The transistor current commutates to the snubber capacitor  $C_{C5}$ . The rate of the rise of drain-source voltage  $U_{T5}$  is slowed by down capacitor  $C_{C5}$ , and thus the zero voltage turn off of the secondary MOSFET is ensured. The maximum value of the voltage  $U_{T5}$  depends on the load current, leakage inductance and the secondary voltage of the high frequency power transformer. While the capacitor  $C_{C5}$  is charging, the capacitor current commutates to the freewheeling diode  $D_0$ .

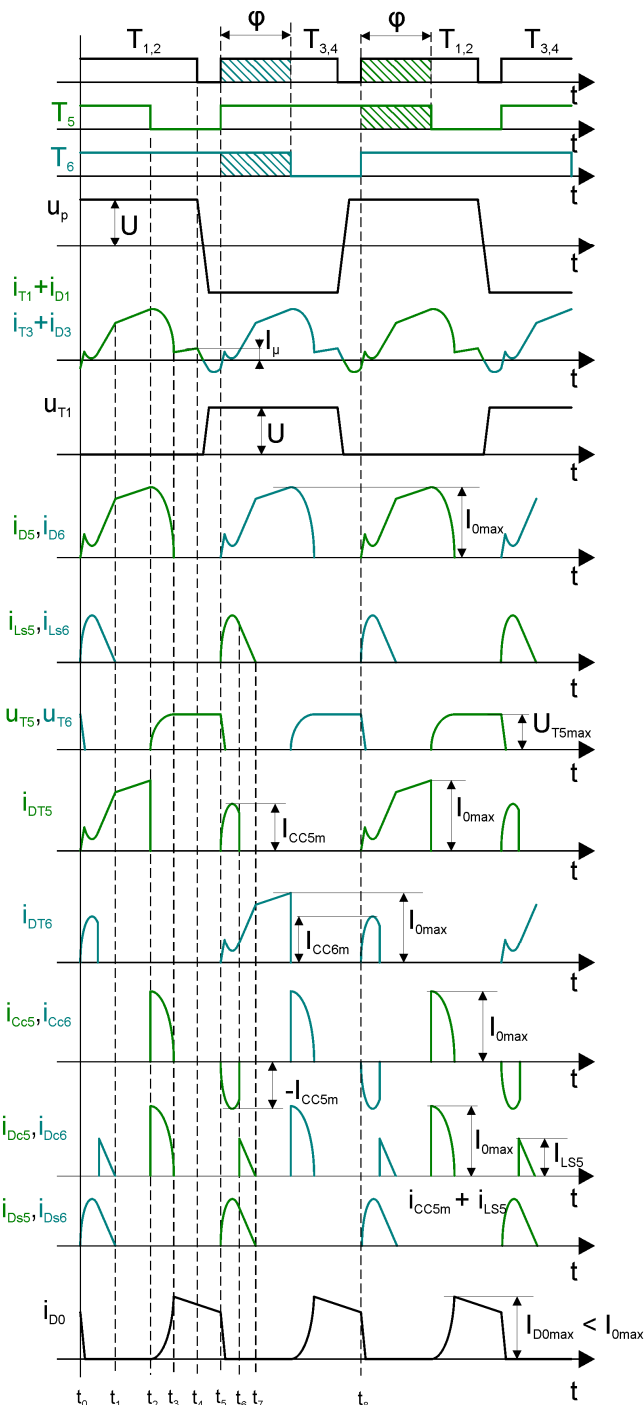


Fig. 2 Operation waveforms of the converter

**Interval ( $t_3$ - $t_4$ ):** Only negligible magnetizing current flows through primary winding of the transformer, so at the time  $t_4$  the primary transistors turn off with negligible losses. The load current starts to flow through the freewheeling diode  $D_0$ .

**Interval ( $t_4$ - $t_5$ ):** The load current continues to flow through the freewheeling diode  $D_0$  during this period.

**Interval ( $t_5$ - $t_6$ ):** The secondary transistor  $T_5$  is turned on at  $t_5$  a half period earlier than primary transistors  $T_1$  and  $T_2$ . The capacitor  $C_{C5}$  starts discharging through  $T_5$ ,  $L_0$ ,  $R_0$ ,  $L_{S5}$ , and  $D_{S5}$ . The rate of rise of discharging current of this capacitor  $C_{C5}$  is limited by the snubber circuit inductance  $L_{S5}$ , and thus zero current conditions of the MOSFET transistor  $T_5$  is achieved. Also the primary transistors  $T_3$ ,  $T_4$  are turned on at

the beginning of this interval and the energy from the input source  $U$  is transferred to the load. But on the beginning of this transfer the energy is supplied also from the snubber circuit capacitance  $C_{C5}$ . The waveforms of the primary and secondary currents are exactly the same like the current waveforms of the opposite transistors only with a half period phase shift. At the time  $t_6$  the discharge current stops to flow through the  $T_5$  transistor.

**Interval ( $t_6$ - $t_7$ ):** The energy stored in snubber inductance  $L_{S5}$  is now flowing through  $D_{S5}$ ,  $D_{C5}$ ,  $L_0$ ,  $R_0$ ,  $L_{S5}$ . At the time  $t_7$  the whole load current flows through the transistor  $T_6$ .

**Interval ( $t_7$ - $t_8$ ):** At the time  $t_8$  ends one period of the DC/DC converter operation and another period starts with the turning on of the transistors  $T_1$ ,  $T_2$ .

#### IV. EXPERIMENTAL RESULTS

A laboratory model of the converter was built to verify the properties of the proposed auxiliary circuit. Following waveforms were obtained at resistive load.

**Fig. 3** shows the waveforms of the primary IGBT transistor (transistor voltage  $u_{CE}$  and current  $i_C$ ), where rate of the collector current rise is slowed down by the leakage inductance of the power transformer. Primary current falls down when the secondary transistors are turned off. IGBT transistor turns off only with negligible turn off losses at the moment when only small magnetizing current flows through primary winding of power transformer. This low turn off and turn on losses conditions for the power IGBT transistors were achieved in full working range from no load to short circuit.

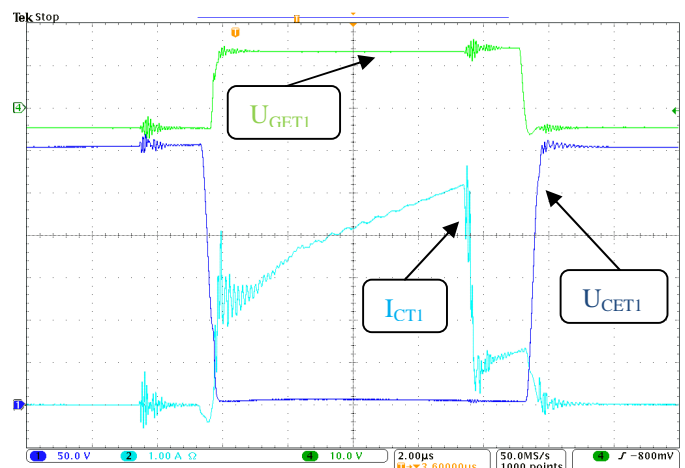


Fig. 3 Primary transistor collector current and collector emitter voltage

**Fig. 4** shows the how the secondary current flowing through MOSFET transistor  $T_5$  commutates to the snubber capacitance  $C_{C5}$  through diode  $D_{C5}$  when the gate source pulse  $U_{GST5}$  is turned off. At the time when the transistor is turned on the rise of the drain current of the  $T_5$  transistor is slowed down by the snubber inductance  $L_{S5}$ .



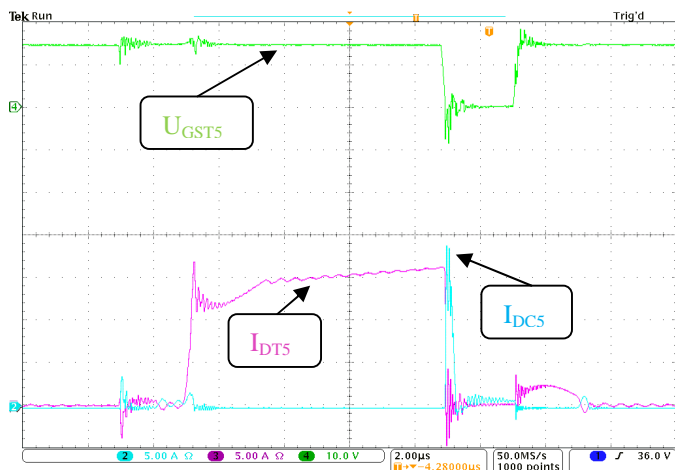


Fig. 4 Secondary transistor drain current commutation

In Fig. 5 the collector-emitter voltage  $u_{DS}$  and collector current  $i_D$  of the secondary MOSFET transistor are shown. The snubber circuit causes a zero voltage turn off of the transistor. The leakage inductance energy of the power transformer obtained at turn off interval is forced to flow through the load at the turn on interval of this transistor. The value of this energy, which should be as low as possible, depends proportionately on the leakage inductance value and with square on value of the transformer current

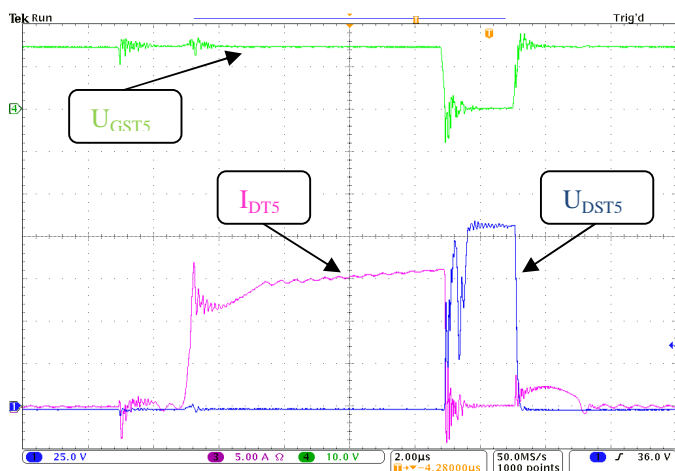


Fig. 5 Secondary transistor drain current and drain source voltage

## V. CONCLUSION

Operation principle of the novel PWM DC/DC converter with secondary snubber is presented in the paper. Soft switching and reduction of circulating currents in the proposed converter are achieved for full load range.

At proper design it is possible to utilize the magnetizing current of power transformer for charging or discharging output capacitances of the IGBT switches and thus zero-voltage turn-on of the IGBTs to achieve. The IGBT transistors are turned-off almost under zero current. Only small magnetizing current of the power transformer is turned-off by IGBT transistors.

The main task of the proposed secondary snubber is transfer of the leakage inductance energy to the load at turn-off of the secondary switch.

Moreover it ensures zero current turn-on and zero voltage turn-off of the secondary switch.

For optimal utilization of the snubber circuit it is necessary to use a power transformer whose leakage inductance is minimized (planar transformer or coaxial transformer). A laboratory model of this converter is being developed to verify the simulations results. The experimental results will be added in the final version of the paper.

A patent application (No. PP 00033-2008) of the proposed new snubber circuit was submitted.

## ACKNOWLEDGMENT

This work was supported by Slovak Research and Development Agency under project APVV-0095-07 and by Scientific Grant Agency of the Ministry of Education of Slovak Republic under the contract VEGA No. 1/0099/09.

## REFERENCES

- [1] Kim E. S., Joe K. Y., Kye M. H., Kim Y. H., and Yoon B. D., "An Improved Soft Switching PWM FB DC-DC Converter for Reducing Conduction Losses," in *Record, IEEE PESC'96*, Vol. I., pp. 651-656.
- [2] Rinne K. H., Thernl K., McCarthy O., "An Improved Zero-Voltage and Zero-Current Switching Full Bridge Converter," in *Record, EPE'95*, Vol. 2., pp. 725-730.
- [3] Lacko M., Olejár M., Ruščin V., Dudrik J.: Converter system for renewable energy utilization with snubber circuit. In: EDPE'07; 16-th international conference on Electrical Drivers and Power Electronics, Proceedings: 24-26 September 2007, The High Tatras, Slovakia; ISBN 978-80-8073-868-6
- [4] Tereň, A., Feňo, I., Špánik, P: DC/DC Converters with Soft (ZVS) Switching. In Conf. Proc. ELEKTRO 2001, section - Electrical Engineering. Žilina 2001, Slovakia, pp. 82 – 90
- [5] Dudrik, J., Špánik, P., Trip, N.-D. : Zero Voltage and Zero Current Switching Full-Bridge DC-DC Converter with Auxiliary Transformer. *IEEE Trans. on Power Electronics*, Vol.21, No.5, 2006, pp. 1328 – 1335.
- [6] Tereň, A., Feňo, I., Špánik, P: DC/DC Converters with Soft (ZVS) Switching. In Conf. Proc. ELEKTRO 2001, section - Electrical Engineering. Žilina 2001, Slovakia, pp. 82 – 90.
- [7] Ruščin V., Olejár M., Lacko M., Dudrik J.: ZVZCS DC-DC converter with controlled output rectifier. In: TRANSCOM 2007 : 7-th european conference of young research and science workers : Proceedings : Žilina June 25-27, 2007. Žilina : University of Žilina, 2007. s. 171-174. ISBN 978-80-8070-694-4
- [8] Lee D. Y., Lee B. K., Hyun D.S.: a novel full-bridge zero voltage transition PWM DC/DC converter with zero-voltage/zero-current switching of auxiliary switches, *PESC 98*, Fukuoka, Japan, pp.961-968
- [9] Dudrik J., Šepeľa J.: Soft-switching current-mode controlled DC-DC converter with secondary switches. In: EDPE 2005 : 13th international conference on Electrical Drives and Power Electronics, September 26-28, 2005, Dubrovnik, Croatia. Zagreb : KoREMA, 2005. 4 p. ISBN 953-6037-43-2
- [10] Dudrik J., Dzurko P.: ARC welder with series-parallel resonant DC-DC converter. In: *Acta Technica CSAV*. vol. 51, no. 4 (2006), p. 415-426. ISSN 0001-7043.
- [11] Lacko M., Olejár M., Ruščin V., Dudrik J.: Non-dissipative turn-off snubber for push-pull converter. In: TRANSCOM 2007 : 7-th european conference of young research and science workers : Proceedings : Žilina June 25-27, 2007. Žilina : University of Žilina, 2007. s. 139-142. ISBN 978-80-8070-694-4..
- [12] Maxim, V., Židek, K., Lupták, M.: Spínač v nule napätia s minimálnym spätným vplyvom na napájaciu sieť. In: *AT&P Journal plus*. č. 1 (2007), s. 163-165. ISSN 1336-5010.

# Shape errors of generators for ADC testing

Michal SAKMÁR

Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

michal.sakmar@tuke.sk

**Abstract**— This paper present some typical errors of exponential stimulus generated by analogue signal sources and their influence on uncertainty of ADC testing by the histogram method. The analyzed exponential signals are very close to linear signals such as triangular and sawtooth signals that can be very simply generated by active integrating circuit or passive integrating circuit with long time constant and/or the final voltage far away from ADC input range. The experimental measurements of some analogue generators on the market as well as generating circuit specially designed for the testing were performed. The limitations were evaluated and the influence of signal errors on uncertainty of testing was determined.

**Keywords**—ADC testing, exponential stimulus signal, capacitor quality.

## I. INTRODUCTION

ADC histogram test methods are widely used and some of them are standardized in IEEE and other standards. One of the main problems of these methods is the requirements of testing signal quality. In [1] the histogram test methods based on exponential stimulus was introduced. The main advantageous of exponential stimulus is that the exponential shape of signal is native shape of any transient effect in electronic circuit. It leads to the idea that exponential signal can be very simply generated and to the expectation that it could be simple to achieve high quality of such signals. In [2] the influence of noise on exponential stimulus histogram test was analyzed. The results showed that the longer is the time constant and the bigger is the final voltage of the signal for the smaller is the influence of noise on the accuracy of testing. The exponential signal meeting these requirements is very close to linear one such as triangular or sawtooth within the input range of ADC under test. The sources of such exponential signals close to linear one are very common - it is generally known that the generation of the most of analogue triangular or sawtooth signals are based on integrating circuit in role of generating circuit, that produce de facto an exponential signal with long time constant and/or with an extreme virtual value of the signal in infinity (such a signal will be called pseudo-linear signal thereafter). This fact leads the author to the idea to employ such analogue generators of pseudo-linear signals for ADC testing by histogram method where, instead of strict linear signal, the exponential stimulus model is used for determination INL and DNL.

The exponential stimulus signal can be described in time domain by next expression:

$$x(t) = (FS + B)e^{\left(\frac{-t}{\tau}\right)} - B, \quad (1)$$

where  $\tau$  is the time constant of the exponential pulse, the interval  $(-FS, FS)$  is the full-scale input range of bipolar ADC under test,  $-B$  is the final value of the exponential signal for  $t \rightarrow \infty$ .

## II. SHAPE ERRORS OF PSEUDO-LINEAR SIGNAL GENERATORS

Some common analogue function generators generating pseudo-linear signal were tested to evaluate to evaluate the restrictions of their application on ADC testing by exponential stimulus histogram method [2]. The output signals of generators were recorded by National Instruments PCI-6289 multifunction card with 18-bit resolution and INL < 10ppm of full scale range [3]. To avoid this residual ADC INL error, the digitized samples were rounded to 16 bits. Then, both exponential and linear fits were calculated using generally known least-square fit method (LMS) to evaluate the applicability of exponential fit and linear fit, respectively. LMS fit method for exponential stimulus (1) leads to the system of nonlinear equations that can not be solved analytically therefore the iteration method by Levenberg-Marquardt algorithm was applied. The differences between the real signals and its LMS fits (shape errors) were calculated in LSB. Because of analogue source of the signals, their do not contain any discontinuities and a relatively small number of samples is needed to estimate their shape error. Therefore the sampling frequency ( $f_s$ ) was chosen so that the total amount of recorded samples within the shot of the signal ( $T_s$ ) over the ADC input range was about 1000 samples. The various combinations of  $T_s$ ,  $f_s$ , ADC input range ( $FS$ ), peak-to-peak value of signal ( $V_{pp}$ ) and DC offset ( $DC$ ) were tested. The figure below shows some typical acquired result. The x-axis indicates sequential number of sample.

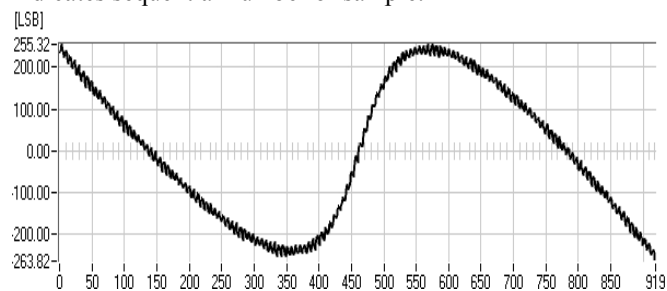


Fig. 1. Generator ESDIGITAL ([4]) FG7002C,  $f_s=300\text{kHz}$  ( $T_s \approx 3\text{ms}$ ),  $V_{pp}=\pm 12\text{V}$ ,  $FS=\pm 10\text{V}$ ,  $DC=0\text{V}$ .

Unfortunately the signal shape errors totally discredit such generators (at least those under test) from accommodation in ADC testing.

### III. SOURCES OF SHAPE ERRORS IN PSEUDO-LINEAR SIGNAL GENERATING CIRCUITS.

To determine the reasons of shape errors of pseudo-linear signals the simple passive and integrating circuit with various types of capacitors, resistors, reference sources and amplifiers have been experimentally evaluated.

First of all the passive integrating circuit was tested according to Fig. 2. NI PCI 6289 was used to generate reference voltage ( $V_{ref}$ ) to charge the capacitor as well as the generated signal digitizer. The recorded samples were rounded to 16bits resolution the same way as within testing of generator hereinbefore.

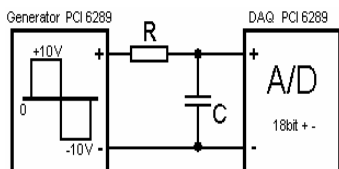


Fig. 2. Test stand for passive integrating circuit

The performed test indicated that the main and the most significant source of signal shape error and the critical component is the capacitor. Some results of testing for the best capacitors are shown in the following figures.

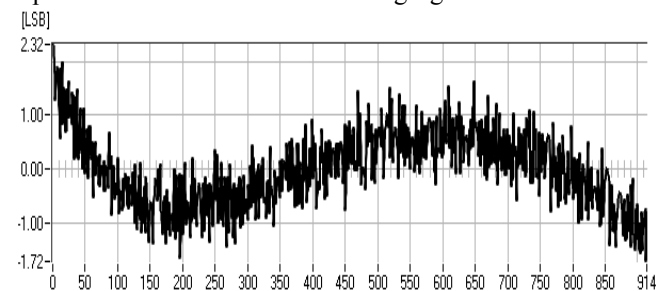


Fig. 3. Capacitor 470n/100V (producer Arcotronics, MKT series: metallised polyester film [5]),  $R = 82k$ ,  $V_{ref} = \pm 10V$ ,  $FS = \pm 5V$

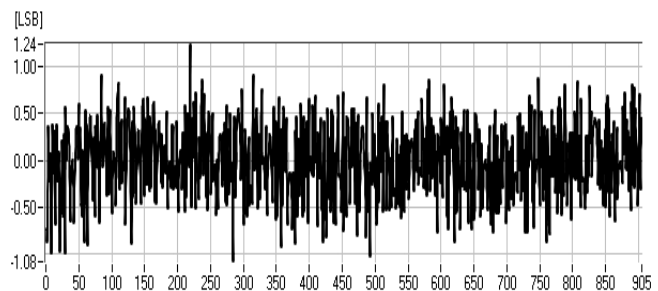


Fig. 4. Capacitor 470n/630V (producer SOLEN, MKP-FC series: metallised Polypropylene film [6]),  $R = 82k$ ,  $V_{ref} = \pm 10V$ ,  $FS = \pm 5V$

The tests indicate (Fig. 3 and 4) that the various types of capacitors (dielectric and other materials, internal construction, etc.) produce different shape errors. On the other hand the performed tests confirmed that some common

capacitors on the market can be accommodated in exponential stimulus generating circuit for testing ADC with at least 15-16 bits resolution.

In Fig. 5 is shown one experimental result acquired the same way as results hereinbefore. It indicates, that convenient operational amplifier can be accommodated in generating circuit without deterioration of signal shape and that the capacitor is the most critical component also here

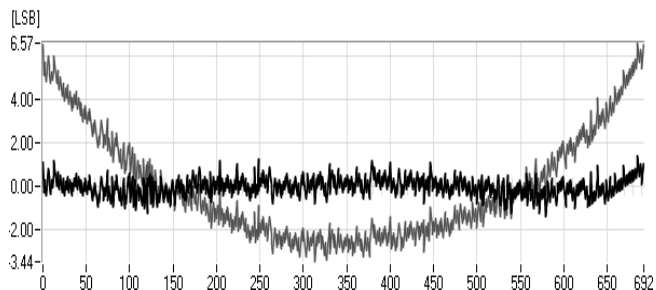


Fig. 5. Signal shape error on the output of active integrator: capacitor SOLEN 470n/630V,  $R=82k$ , amplifier: Analog Devices OP284F ([7], power  $\pm 12V$ ,  $V_{pp}$  on output  $\pm 6V$ ),  $FS = \pm 5V$ , exp. (black) and linear (grey) fits.

### IV. CONCLUSION

The paper introduces simple analysis of deterioration of exponential stimulus for ADC histogram test method based on experimental measurements. The experiments show that the common analogue function generators are not applicable for ADC testing even using the exponential fit. The most critical component in practical realization of stimulus generating circuit is the capacitor. On the other hand, some common capacitors on the market enable realization an accurate passive and/or active generating circuit. Besides a simple model of signal deterioration was suggested. The model allows simple estimating the limitation of a generator for required ADC testing by exponential histogram method.

### ACKNOWLEDGMENT

The work is a part of project supported by the Science Grant Agency of Slovak republic (No. 1/0103/08).

### REFERENCES

- [1] Holcer, R. – Michaeli, L. – Šaliga, J.: *DNL ADC testing by the exponential shaped voltage*, IEEE transactions on instrumentation and measurement, ISSN 0018-9456, Vol. 52, no. 3 (2003), p. 946-949.
- [2] Šaliga, J. – Michaeli, L. – Holcer, R.: *Noise sensitivity of the exponential histogram ADC test*, Measurement, ISSN 0263-2241, Vol. 39, no. 3 (2006), p. 238-244
- [3] www.ni.com
- [4] <http://www.testequipmentdepot.com/ezdigital/fg7002c.htm>
- [5] [http://www.arcotronics.it/pdf/general\\_short\\_form.pdf](http://www.arcotronics.it/pdf/general_short_form.pdf)
- [6] www.solen.ca
- [7] www.analog.com.

# Thermal degradation of transformer oils

<sup>1</sup>Peter SEMANČÍK

<sup>1</sup>Dept. of Electric Power Engineering, FEI TU of Košice, Slovak Republic

<sup>1</sup>peter.semancik@tuke.sk

**Abstract**—This paper deals with the diagnostic method thermal – oxidation stability of transformer oils. It describes the principle test method as a preparation of experiment. Three samples were used for the transformer oil experiment.

**Keywords**—Oxidation stability, oil life, dielectric dissipation factor, ČEZ – ORGREZ test method, transformer oil, diagnostics.

## I. INTRODUCTION

Among material used in electrical transformers, transformer oil has special state. It is not only for its extraction, but also for several use like insulation or coolant.

Insulating oils should have stable high-quality properties, not only in the original state, but also during the up time in operation. The stability of insulating oils has an elementary meaning during operation, because they work under high temperatures usually by the presence of oxygen, so they should be oxidation resistant.

Phenomena, which controlled natural oil toward change, chemical alternatively electric facilities in working condition, and which they can subject cut-down safety in service facilities, are called ageing process.

Speed ageing process of transformer oil can be affected by temperature (growth temperature about 8 till 10K will cause doubling velocity ageing process) and catalyzer (especially copper and iron).

There are several diagnostic methods, which deal with actual problems of insulating transformer oils.

The properties of insulating oil and measuring methods e.g. (STN EN 60296):

- viscosity (ISO 3104, IEC 61868),
- water content (IEC 60814),
- dielectric dissipation factor (DDF), (IEC 60247, IEC 61620 at 90 °C),
- oxidation stability (IEC 61125, method C),
- antioxidant additive content (IEC 60666),
- interfacial tension (IFT), (ISO 6295),
- breakdown voltage (IEC 60156).

This paper describes two of the test methods. Diagnostic methods are thermal – oxidation stability of insulating fluids by the ČEZ– ORGREZ test method [1].

## II. ČEZ - ORGREZ METHODOLOGY

Oxidation stability is an indicator that allows us to set stricter limits for oils in special applications. In some countries, stricter limits or other requirements and tests are imposed.

During the test, the sample of new or reclaimed oil is exposed to conditions simulating a load application, similar to the load during operation. Individual factors are simplified. High-quality parameters are periodically monitored until sediments are formed and the oil is no longer usable.

### *Preparation of the experiment, and testing process [1]*

Before the test begins, the initial high-quality parameters are determined and 500ml is taken away from the oil sample.

A measuring cylinder is used to measure out 5000 ml of oil, marked out for testing, into the sulphonation flask.

The required quantity of copper wire will be added to the oil – 10g (quantity cca. 0,1 cm<sup>2</sup>/g of oil) to each liter of oil.

The flask with oil will be put into the laboratory drying chamber, at a temperature of 100 °C.

Using tubes for example made of glass, air will be conducted into the samples to ensure delivery of condensed fluid into the separation trap outside the drying chamber.

A control test and verification of the temperature regulated in the drying chamber will be carried out every day.

Some, of the samples, will be removed at weekly intervals to determine the values of selected parameters (acidity, interfacial tension and content of inhibitors – 1x 168 hours, dielectric dissipation factor – 1x per 336 hours).

The test will be completed when sediments insoluble in n-heptane are present or when there are no more samples for continuing the test or after 840 hours of testing.

## III. EXPERIMENT

The thermal – oxidation stability test of insulating oil was made using the ČEZ – ORGREZ method [1].

Three samples of power transformer insulating oil were used.

Further information about the sample is confidential to the manufacturer and to the plant operator. During the experiment, the data, interpreted in (Table) were measured [4].

The principle of the test is based on the air oxidation of the measured oil with added accelerator at a given temperature.

TABLE I  
VALUES MEASURED BY THE ČEZ – ORGREZ TEST METHOD [4]. FOR SAMPLE NO. 3

Test period (h)	tgδ (x10 <sup>-2</sup> )			ε <sub>r</sub> (-)			ρ (GΩ m)			ČK	σ	Q <sub>i</sub>	Sediments insoluble in the n-heptans
	20°C	70°C	90°C	20°C	70°C	90°C	20°C	70°C	90°C				
0	0,001	0,003	0,008	2,200	2,125	2,094	65789,4	59210,5	25000,0	0,003	58	0,38	-
168										0,004	56	0,39	-
336	0,009	0,050	0,063	2,184	2,136	2,104	25165,0	13228,0	1937,0	0,007	55	0,39	-
504										0,007	55	0,38	-
672										0,007	55	0,38	-
840	0,002	0,007	0,015	2,203	2,146	2,122	6788,0	2819,6	1730,2	0,004	54	0,38	-
1008										0,004	54	0,36	-
1176										0,005	52	0,33	-
1344	0,007	0,012	0,023	2,196	2,136	2,105	4602,0	1260,2	945,1	0,004	52	0,33	-
1512										0,008	51	0,30	-
1680										0,008	51	0,28	-
1848	0,006	0,028	0,063	2,201	2,131	2,100	1429,1	714,5	321,5	0,007	51	0,25	-
2016										0,009	50	0,22	-
2184										0,008	49	0,21	-
2352	0,088	0,112	0,129	2,187	2,141	2,113	1421,2	387,6	193,8	0,008	47	0,21	-
2520										0,008	48	0,16	-
2688										0,009	47	0,15	-
2856	0,082	0,136	0,145	2,199	2,156	2,112	1503,5	294,1	130,7	0,010	47	0,11	-
3024										0,007	47	0,12	-
3192										0,004	46	0,08	-
3360	0,158	0,293	0,311	2,196	2,127	2,099	1216,6	425,8	121,6	0,007	46	0,08	-

ČK – mg KOH/g,

σ – mN/m,

Q<sub>i</sub> – % wt.,

– sample filtered using white tape filter

paper (6 – 6,8 μm) before the measurement. 90 °C). The dependence of tgδ rises most at the temperature 90 °C in comparison with temperature 20 °C.

The test was carried out under the following conditions [1]:

- temperature 100 °C,
- volume of oil samples 5 l,
- bubbling of oil dried and refined by air in larger amounts than are needed for reaction of oil with the air,
- accelerator: copper wires in quantities cca 0,1 cm<sup>2</sup>/g measured oil.

The separation trap, placed outside the drying chamber, gathers the condensed fluid released during the test.

The values measured in dependency on the length of test periods are recorded in tables (Table 1), which show the degradation process of the oil until the moment when sediments insoluble in n-heptane form or until the test is terminated.

Monitored parameters [2-5]:

tgδ – dielectric dissipation factor,

ε<sub>r</sub> – dielectric permittivity,

ρ – volume resistivity,

ČK – determination of acidity,

σ – determination of interfacial tension of oil against water,

Q<sub>i</sub> – contents of inhibitors.

The graphic dependencies in Fig. 1 – Fig. 6 were made from the measured values monitoring the individual parameters.

Figure 1 shows the dielectric dissipation factor as a function tgδ=f(t) at different temperatures (20 °C, 70 °C,

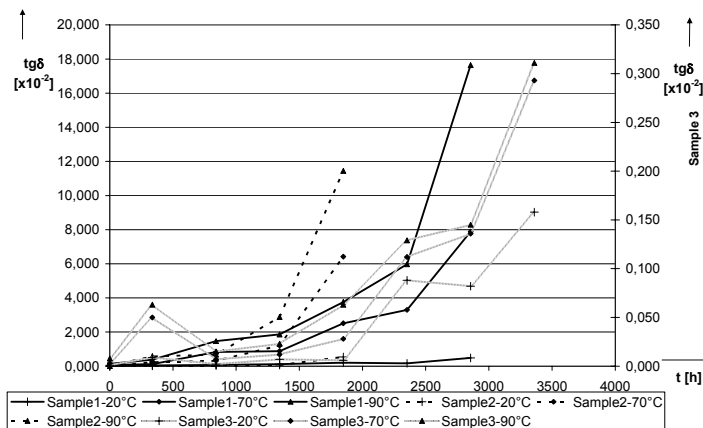


Fig. 1. Dependence of tgδ. Samples 1 – 3.

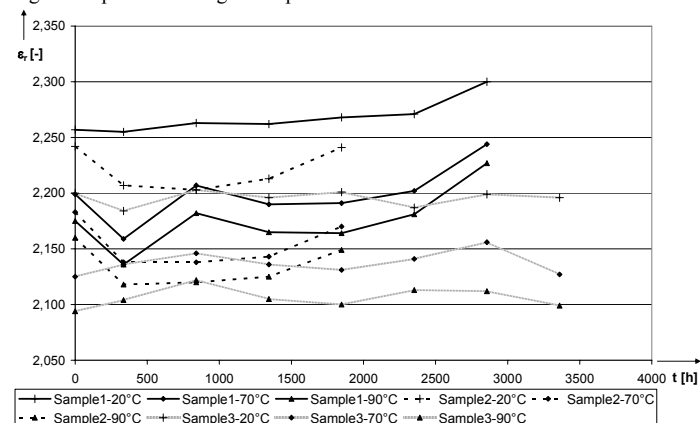


Fig. 2. Dependence of ε<sub>r</sub>. Samples 1 – 3.

Figure 2 shows the dielectric permittivity as a function  $\epsilon_r=f(t)$  at different temperatures (20 °C, 70 °C, 90 °C). The dependence of  $\epsilon_r$  changes most at the temperature 20 °C in comparison with the temperature 90 °C. When comparing  $\text{tg}\delta(90\text{ °C})$  with  $\epsilon_r(90\text{ °C})$  we can see that the dependence of  $\text{tg}\delta(90\text{ °C})$  rises and dependence of  $\epsilon_r(90\text{ °C})$  changes minimally.

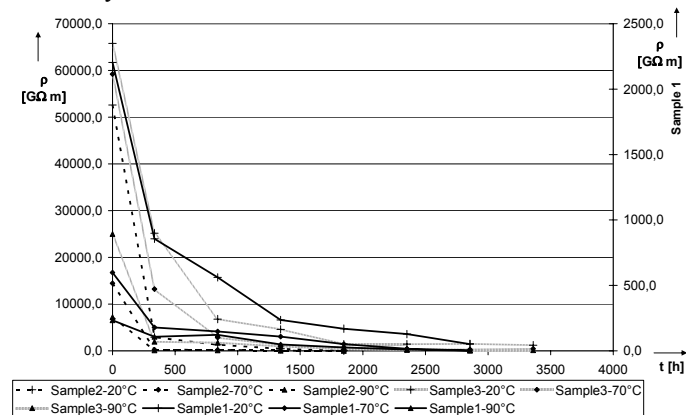


Fig. 3. Dependence of  $\rho$ . Samples 1 – 3.

Figure 3 shows the volume resistivity as a function  $\rho=f(t)$  at different temperatures (20 °C, 70 °C, 90 °C).

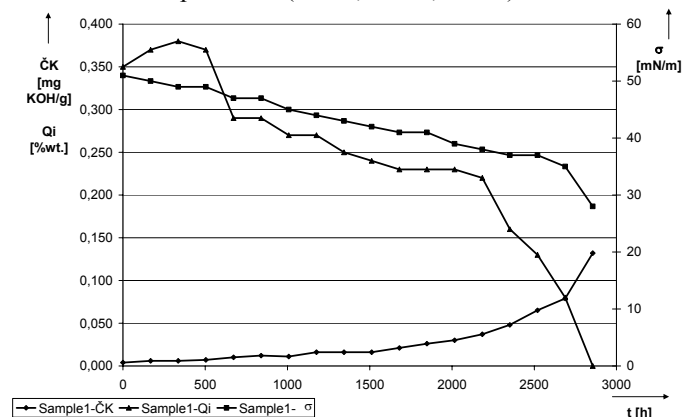


Fig. 4. Dependence of  $\check{C}K$ ,  $Q_i$ ,  $\sigma$ . Sample 1.

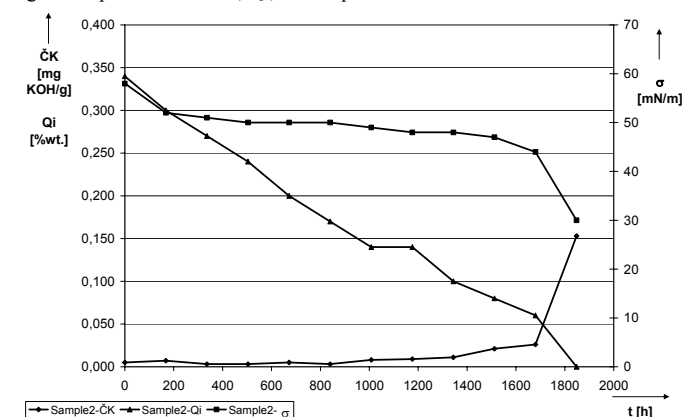


Fig. 5. Dependence of  $\check{C}K$ ,  $Q_i$ ,  $\sigma$ . Sample 2.

Figures 4 - 6 shows the  $\check{C}K$ ,  $Q_i$ ,  $\sigma$  as a function  $(\check{C}K, Q_i, \sigma)=f(t)$ . The dependence of  $\check{C}K$  lightly rises and dependence of  $\sigma$  lightly decreases until the creation of sediments (that are insoluble in insulating oil), when process rapidly changes (growths, decreases). Dependence of  $Q_i$  decreases equally to the duration of sediments creation, which is insoluble in insulating oil.

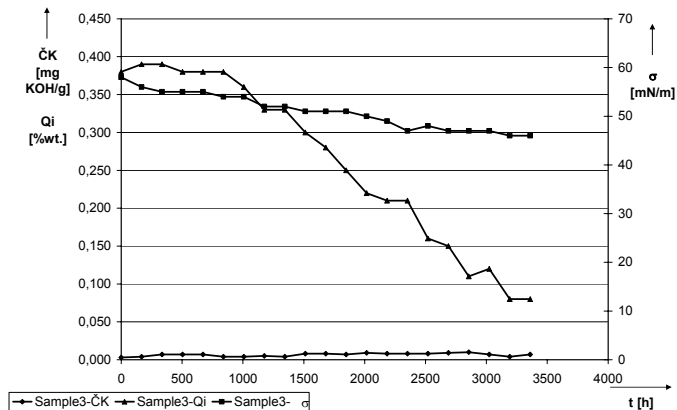


Fig. 6. Dependence of  $\check{C}K$ ,  $Q_i$ ,  $\sigma$ . Sample 3.

The graphic dependencies show degradation process of the oil until the moment when sediments insoluble in n-heptane form. Until the time of insulating oil degradation (Sample 2  $\approx$  1400 h, Sample 1  $\approx$  2000 h) the dependencies of  $\text{tg}\delta$ ,  $\epsilon_r$ ,  $\check{C}K$ ,  $\sigma$  are changing minimally. In the next degradation process until the duration of sediments creation, which are insoluble in insulating oil, are the dependencies and characteristic of insulating oil changing rapidly. The dependence of  $Q_i$  decreases equally to the duration of sediments creation, which is insoluble in insulating oil. For the diagnostics of insulating oil degradation process, monitoring the parameter  $Q_i$  is necessary, because the dependence and gradient of the characteristics are changing equally during the whole degradation process. From this it is possible to follow the change of insulating oil properties, not only at the end of degradation process, but also at the beginning, in comparison with the other monitored parameters. During the degradation process of sample No.3 (Sample 3 = 3360 h) did not come to the creation of sediments, which are insoluble in n-heptans, because there were no more samples for continuing the test.

#### IV. CONCLUSION

The oxidation stability of oil is evaluated by period of time until sediments that are soluble in the insulating oil (insoluble in the n-heptane), or by the creation of sediments that are insoluble in insulating oil. In the test of thermal-oxidation stability the submitted 3 samples of insulating oil degraded for sample 1 in 2856 hours, sample 2 - 1848 hours, sample 3 – 3360 hours. This was documented by the presence of sediments insoluble in the n-heptanes except of sample 3. The thermal – oxidation stability test was carried out using, the ČEZ – ORGREZ method.

#### ACKNOWLEDGMENT

This work was supported by scientific Grant Agency of the ministry of Education of the Slovak Republic project VEGA No. 1/0368/09 and APVV-20-006005.

#### REFERENCES

- [1] SOP 2-32/72: The thermal – oxidation stability of insulating fluids by the CEZ ORGREZ test method, ORGREZ a.s., Brno, Czech Republic, 2004.
- [2] STN EN 60296: Fluids for electrotechnical applications. Unused mineral insulating oils for transformers and switchgear. Slovak standard Institute, Bratislava, 2005.
- [3] ČSN EN 61125: Unused hydrocarbon-based insulating liquids. Test methods for evaluating the oxidation stability. Czech standards institute, 1996.
- [4] ORGREZ, a.s.: Protocol about measuring. Experiment of the insulating oil thermal – oxidation stability CEZ-ORGREZ method. ORGREZ a.s. Electrical Engineering Laboratory Division, Praha, 2006.
- [5] CEZ, a.s.: Prophylactic of mineral insulating oils. Company standard 00/08 rev0.

# Comparison of interpolation methods for estimation of Rayleigh fading channel in OFDM system

Ján ŠTERBA

Department of Electronics and Multimedia Communications, Faculty of Electrical Engineering and Informatics,  
Technical University of Košice, Slovak Republic

sterba.jan@gmail.com

**Abstract**—In this paper, pilot symbol assisted channel estimation in OFDM system via different interpolation methods is presented. In order to perform coherent detection of transmitted data, reliable estimation of channel is required. This can be obtained by occasionally transmitting known data, or so called "pilot symbols" and perform filtration or interpolation on these pilot symbols scattered in an OFDM. In this paper, pilot symbol interpolation via Bezier curve interpolation is examined, and the performance of this method of channel information interpolation in pilot symbol assisted OFDM system in frequency selective multipath Rayleigh fading channel is evaluated by measuring the bit error rate (BER). The results are then compared with linear interpolation, spline interpolation, cubic interpolation, Lagrange barycentric interpolation and perfect channel state information.

**Keywords**—OFDM, channel estimation, Rayleigh fading channel, pilot symbols, FFT interpolation, Cubic interpolation, Lagrange interpolation.

## I. INTRODUCTION

Orthogonal Frequency Division Multiplexing (OFDM) is a very promising technology due to its high data rate transmission capability, high bandwidth efficiency and its robustness to multipath delay. It is a very potential technique for transmitting high-bit-rate data over indoor and outdoor wireless communication channels, and it has been already used in several broadcast systems [1] and wireless LAN standards [2].

Since the radio channel is frequency selective and time-varying for wideband mobile communications systems [3], in order to perform coherent detection, reliable channel estimates are required. These can be obtained by occasionally transmitting known data or so called "pilot symbols". As the pattern and parameters of pilot symbols are known both by the receiver and the transmitter, the channel gain and phase distortion can be easily obtained from the received signal at the receiver side at the pilot symbol positions. The filtration or interpolation of channel information between pilots is then necessary to perform to obtain the channel estimation for the data signals.

In [4], [5] and [6], different filtering algorithms have been proposed for pilot symbol assisted channel estimation. The minimum mean-square error (MMSE) channel estimation for block pilots was proposed in [7] and the 2-D optimum

Wiener filtering for pilot symbol assisted OFDM channel estimation was studied in [8]. However, the filtering algorithms for pilot symbol assisted channel estimation require channel statistics, such as delay profile and the Doppler frequency, which are usually unknown in typical mobile wireless situations [9]. In the slow fading channels, the channel information can be further enhanced by decision directed channel estimation. The channel estimation can be also realized utilizing interpolation methods on pilot symbols. Many different types of interpolation can be used, like the linear interpolation, spline interpolation, FFT interpolation, cubic interpolation and others. In this paper, our aim is to describe 2-D FFT interpolation and compare its results with other interpolation methods in frequency selective multipath Rayleigh fading channel with 8-QAM (Quadrature Amplitude Modulation) modulation scheme.

Different pilot patterns have been also studied in the literature. Comparison study of five types of pilot patterns is presented in [10]. In the literature, two types of pilot patterns were widely studied – the block pilots and comb pilots [11]. The block type pattern performs best in slow varying channels, and the comb pilots usually perform better than block type pilot pattern in fast fading channels in tracking the time variation of the channel. However, the rectangular arrangement is the one being used in many practical applications, and for this reason has been chosen for our simulations.

The rest of the paper is organized as follows. Section II describes OFDM system model as being used in simulations. Section III presents multipath Rayleigh fading channel. Section IV describes interpolation techniques of channel state information and section V summarizes simulation results and presents comparison study of different interpolation methods. Finally, section VI contains the conclusion.

## II. OFDM SYSTEM

The OFDM system used in simulations is shown in Fig. 1. The digital binary information is first mapped to modulation in 'signal mapper' block. After signal is converted from serial to parallel form, pilot symbols are inserted according to rectangular pattern. Following the conversion, Inverse Discrete Fourier Transform (IDFT), is used to transform sequence of data with length  $N\{X\}$  into time domain

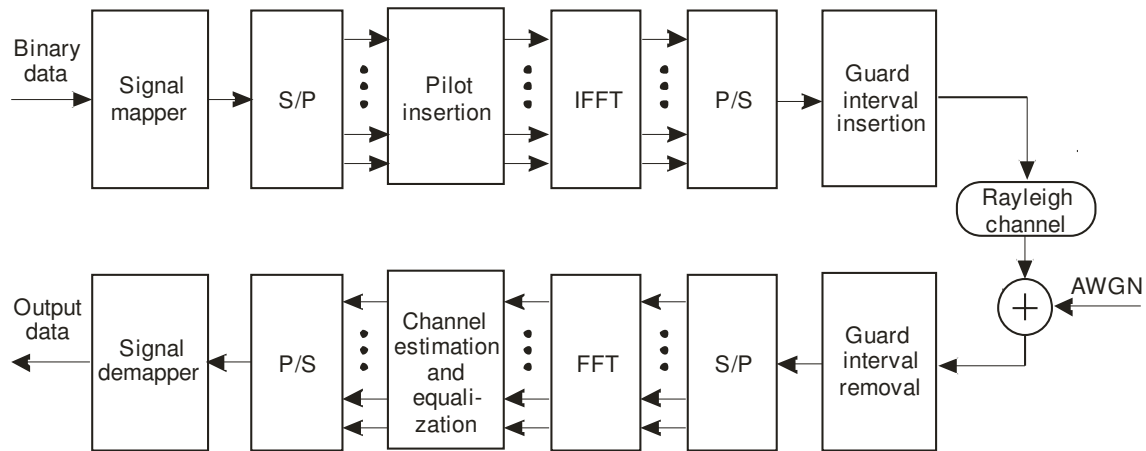


Fig. 1. Baseband OFDM system model with pilot symbol insertion.

signal  $\{x(n)\}$  according to:

$$\begin{aligned} x(n) &= IDFT\{X(k)\} \quad n = 0, 1, 2, \dots, N-1 \\ &= \sum_{k=0}^{N-1} X(k) e^{j(2\pi kn/N)} \end{aligned} \quad (1)$$

where  $N$  is length of DFT. The guard time, which consists of cyclically extended part of OFDM symbol, is then chosen larger than expected delay spread, and inserted to prevent inter-symbol interference (ICI). The resulting OFDM symbol is given as:

$$x_f(n) = \begin{cases} x(N+n), & n = -N_g, -N_g+1, \dots, -1 \\ x(n), & n = 0, 1, \dots, N-1 \end{cases} \quad (2)$$

where  $N_g$  is the length of guard interval. The resulting signal  $x_f(n)$  will be finally transmitted through frequency selective multipath Rayleigh fading channel, which is shown in Fig. 2, and additive noise will be added. The received signal is given by:

$$y_f(n) = x_f(n) \otimes h_f(n) + w(n)$$

where  $w(n)$  is additive white Gaussian noise (AWGN) and  $h_f(n)$  is channel impulse response. At the receiver, the guard time is removed and signal is sent to DFT block for following operation:

$$\begin{aligned} Y(k) &= DFT\{y(n)\} \quad k = 0, 1, 2, \dots, N-1 \\ &= \frac{1}{N} \sum_{n=0}^{N-1} y(n) e^{-j(2\pi kn/N)}. \end{aligned} \quad (3)$$

Then, the pilot symbols are extracted and the channel is estimated in the channel estimation block. The transmitted data is then estimated by following operation:

$$X_c = \frac{Y(k)}{H_c(k)} \quad k = 0, 1, \dots, N-1 \quad (4)$$

where  $H_c(k)$  is the estimated channel obtained by channel estimation block. Then, the binary information data is obtained back in 'signal demapper' block.

### III. RAYLEIGH FADING CHANNEL

When a continuous waveform is transmitted through such a channel, that the multipath effect results in the fluctuation of the received signal envelope which is Rayleigh distributed, the transmission channel is known as a Rayleigh fading channel. The time variations of the signal level is characterized by the Doppler frequency effect, which is caused by the motion of

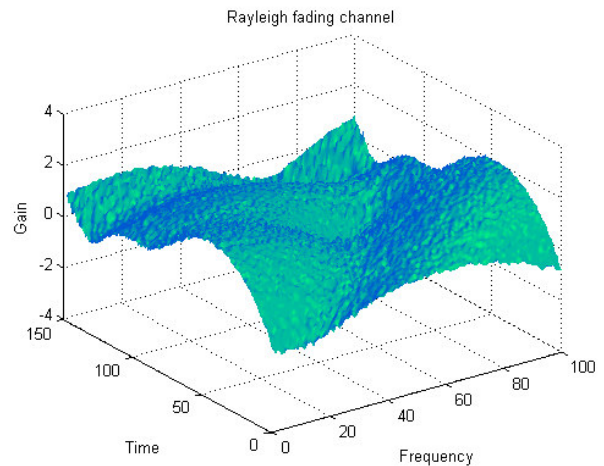


Fig. 2. Time-varying frequency characteristic of frequency selective multipath Rayleigh fading channel, as employed in simulations.

the mobile terminal. Let  $A$  denote a received signal to noise ratio, which is proportional to the square of the signal envelope. The p.d.f of  $A$  is exponential and can be written as [12]:

$$p_A(a) = \frac{1}{\rho} \exp\left\{-\frac{a}{\rho}\right\}, \quad \rho = E[A] \quad (5)$$

for  $a \geq 0$ .

After derivation in, let  $f_m$  be the maximum Doppler frequency defined as

$$f_m = \frac{v}{\lambda}, \quad (6)$$



where  $\mathbf{v}$  is the speed of the terminal and  $\lambda$  is the wavelength. Then, the expected number of times per second  $N_a$  that received SNR will pass downward a given level  $a$  is:

$$N_a = \sqrt{\frac{2\pi a}{\rho}} f_m \exp\left\{-\frac{a}{\rho}\right\}. \quad (7)$$

#### IV. INTERPOLATION TECHNIQUES

The FFT based interpolation is a high-resolution interpolation based on zero-padding and consecutive FFT and IFFT. After obtaining the estimated channel  $\{H_p(k), k = 0, 1, \dots, N_p - 1\}$  it is first necessary to convert it to time domain by IFFT:

$$G(n) = \sum_{k=0}^{N_p-1} H_p e^{j(2\pi kn/N_p)}, \quad n = 0, 1, \dots, N_p - 1 \quad (8)$$

Then, by using the basic multi-rate signal processing properties, the signal is interpolated by transforming  $N_p$  points into  $N$  points with the following method:

$$M = \frac{N_p}{2} + 1$$

$$G_N(n) = \begin{cases} G_N, & 0 \leq n < M - 2 \\ 0, & \frac{N_p}{2} \leq n - M \\ G_N(n - N + 2M - 1), & -M \leq n - N < -1 \end{cases} \quad (9)$$

The estimate of the channel at all frequencies is then obtained by:

$$H(k) = \sum_{n=0}^{N-1} G_n(n) e^{-j(2\pi/N)nk}, \quad 0 < k \leq N - 1. \quad (10)$$

The 1-dimensional FFT interpolation method, as is illustrated in Fig. 3, takes as input a block of  $N_p$  samples,  $x(nT)$ ,  $n = 0, \dots, N - 1$ , and outputs a block of  $MN$  samples,  $y(mT/M)$ ,  $m = 0, \dots, MN - 1$ . The output has the required property of ideal interpolator that  $y(nT) = x(nT)$ . With this approach, the channel is first estimated in frequency domain on the  $N_{pilots}$  sub-carriers where pilots have been transmitted on. Then, this

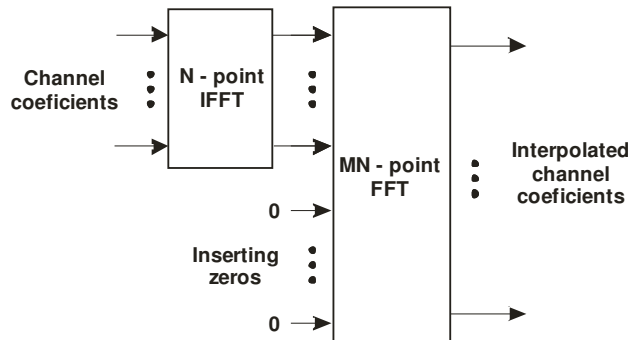


Fig. 3. Low complexity FFT based channel estimator.

$N_{pilots}$  estimates are transformed with an  $N_{pilots}$  point IFFT in the time domain and the resulting time sequence is transformed back in the frequency domain with  $MN$  point FFT, where  $MN$  was created by adding zeros to  $N_{pilots}$ , and  $MN > N_{pilots}$ .

The FFT interpolation can also be applied in 2-dimensions, with interpolation in the time and frequency directions, by successive FFT [13]. An example of 2-D interpolation of Rayleigh fading channel can be seen in Fig. 4, which shows time-varying frequency characteristic of channel. The upper figure shows the channel information gathered from pilot symbols, and the lower figure shows its interpolated associate.

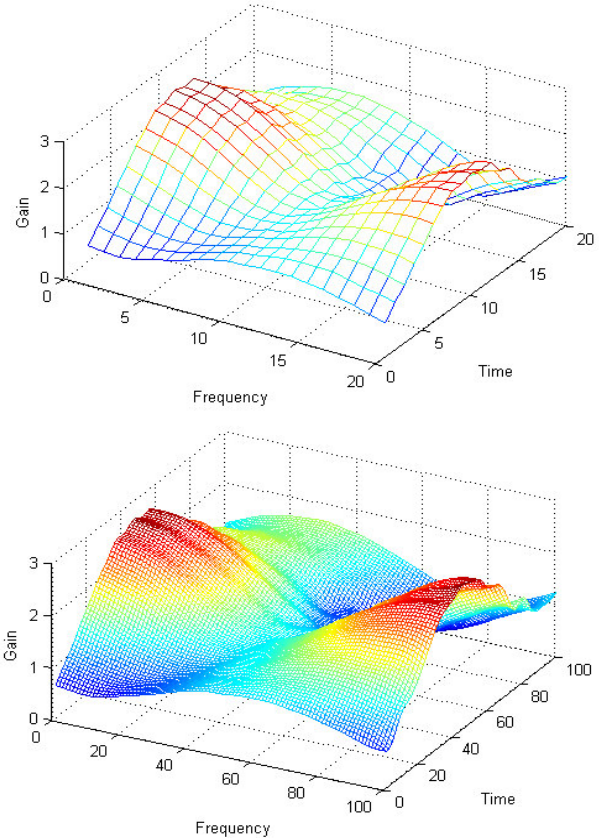


Fig. 4. 2-Dimensional FFT based interpolation of time-varying frequency characteristic of Rayleigh fading channel, as performed on pilot symbols.

Now the other interpolation methods will be described shortly. The linear interpolation is shown to perform better than the piecewise-constant interpolation. The channel estimation at the data subcarrier  $k$ ,  $mL < k < (m+1)L$ , using linear interpolation is given as:

$$H_e(k) = H_e(mL + l) \quad 0 \leq l < L$$

$$= H_p(m+1) - H_p(m) \frac{l}{L} + H_p(m). \quad (11)$$

The Bezier curves are usually used for smoothing, but might be as well used for interpolation. Bezier curve can be created from channel estimation at data subcarriers by:

$$B_C(t) = (1-t)^3 H_p(0) + 3(1-t)^2 t H_p(1) + 3(1-t)t^2 H_p(2) + t^3 H_p(3), \quad t \in [0, 1]. \quad (12)$$

The estimate of the channel at data subcarriers can than be obtained by:

$$H_e(k) = B_C(t), \quad \text{where } t = k.$$

To interpolate between pilot points using spline interpolation, an individual piecewise polynomial is constructed between each segment, e.g.:

$$S(x) = \begin{cases} S_0(x) & x \in [x_0, x_1] \\ S_1(x) & x \in [x_1, x_2] \\ \vdots & \vdots \\ S_{n-1}(x) & x \in [x_{n-1}, x_n] \end{cases} \quad (13)$$

where  $x_i$  is a position of pilot symbol and each  $S_i(x)$  is a unique spline interpolant, which can be found using following equation:

$$S_i(x) = H_{p,i} + z_i(x - x_i) + \frac{z_{i+1} - z_i}{s(x_{i+1} - x_i)}(x - x_i)^2 \quad (14)$$

where  $z_{i+1} = -z_i + 2 \frac{H_{p,i+1} - H_{p,i}}{x_{i+1} - x_i}$ .

### V. RESULTS

In this section, the performance of different interpolation methods for pilot symbol assisted channel estimation is demonstrated. In the simulations, the total number of used subcarriers was 128, with FFT size 1024 and pilot symbol ratio 1/12. As channel model was used Rayleigh fading channel with Doppler frequency 50 Hz. The resulting bit error rate for different interpolations is presented in Fig. 5. The bit error rate is degraded both by imperfect channel estimates and noise disturbances. Finally, the resulting BER curves are compared with BER computed with perfect channel state information. As can be seen from the figure, the FFT interpolation performs the best from all other compared methods.

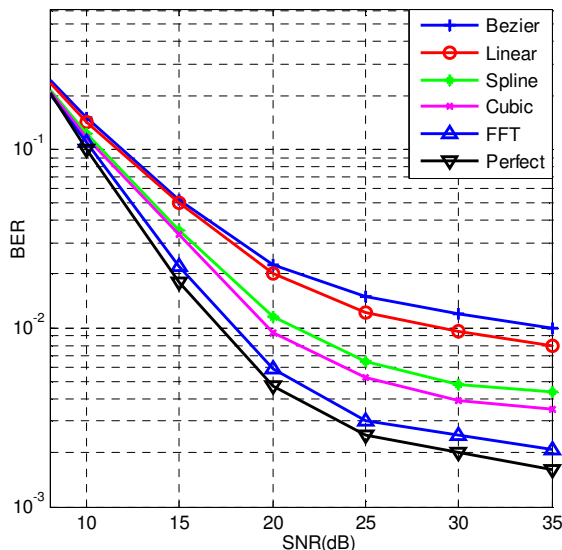


Fig. 5. Bit error rate at different signal to noise ratios for 8-QAM in frequency selective Rayleigh fading channel with Doppler freq. 50 Hz.

### VI. CONCLUSION

In this paper, different interpolation methods of channel information have been investigated in pilot symbol assisted OFDM system. The bit error rate has been calculated and compared with different interpolation methods, like cubic interpolation, linear interpolation, spline interpolation and 2-D FFT based interpolation. The simulation results show that pilot symbol assisted channel estimation with rectangular grid performs best with FFT interpolation among all compared channel interpolation algorithms. This estimation method can be used to efficiently estimate the channel in an OFDM system without any knowledge of channel statistics, and is highly robust against channel degradation.

### ACKNOWLEDGMENT

This work has been funded by the Grant Agency SPP Hlavička, and also funded by VEGA 1/4088/07 "Rekonfigurovatelné platformy pre širokopásmové bezdrôtové telekomunikačné siete" and COST 297: "High Altitude Platforms for Communications and other Services".

### REFERENCES

- [1] ETSI EN 300 744 v1.5.1, Digital Video Broadcasting (DVB); Framing structure, channel coding and modulation for digital terrestrial television. Nov. 2004.
- [2] Hui Li, G. Malmgren, M. Pauli, "Performance comparison of the radio link protocols of IEEE802.11a and HIPERLAN/2." Vehicular Technology Conference, Sept. 2000 p. 2185 - 2191.
- [3] A. R. S. Bahai and B. R. Saltzberg, Multi-Carrier Digital Communications: Theory and Applications of OFDM. Kluwer Academic/Plenum, 1999.
- [4] G. Auer, E. Karapidis, "Pilot aided channel estimation for OFDM: a separated approach for smoothing and interpolation." IEEE International Conference on Communications, 2005.
- [5] J.V. Choi and Y. H. Lee, "Design of 2-D channel estimation filters for OFDM systems," IEEE Intl. Commun. (ICC'05), p. 2568-2572, May 2005.
- [6] H.H. H'mimy, "Channel estimation based on coded pilot for OFDM." Vehicular Technology Conference, 1997 IEEE 47th. Volume 3, p. 1375 - 1379, 1997.
- [7] J.-J. Van de Beek, O. Edfords, M. Sandell, S.K. Wilson and P.O. B'orjesson, "On channel estimation in OFDM systems." Proc. IEEE Vehicular Technology Conference (VTC'95), vol. 2, Chicago, IL, p. 815-819, July 1995.
- [8] P. Hoeher, S. Kaiser and P. Robertson, "Two-dimensional pilot-symbol-aided channel estimation by Wiener filtering." IEEE Intl. Conf. Acoustics, Speech, and Signal Processing (ICASSP'97), Munich, Germany, p. 1845-1848, Apr. 1997.
- [9] Ye Li. Pilot-symbol-aided channel estimation for OFDM in wireless systems. IEEE Transactions on Vehicular Technology, Volume 49, Issue 4, p. 1207 - 1215, July 2000.
- [10] F. Tufvesson and T. Maseng, "Pilot assisted channel estimation for OFDM in mobile cellular systems." Proc. IEEE VTC'97, p. 1639-1643, May 1997.
- [11] Xiaodai Dong, Wu-Sheng Lu, A.C.K. Soong, "Linear Interpolation in Pilot Symbol Assisted Channel Estimation for OFDM." Wireless Communications, IEEE Transactions on Volume 6, Issue 5, p. 1910 - 1920, May 2007.
- [12] H. S. Wang, N. Moayeri, "Modeling, Capacity, and Joint Source / Channel Coding for Rayleigh Fading Channels", Vehicular Technology Conference, IEEE 43<sup>rd</sup>, 1993.
- [13] J. W. Adams, "A subsequence approach to interpolation using the FFT," IEEE Trans. Circuits Syst., vol. CAS-34, pp. 568-570, May 1987.

# Taylor Series Based Tracking Algorithm

Mária ŠVECOVÁ

Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

Maria.Svecova@tuke.sk

**Abstract**—Target tracking by using measurement of time of arrival belongs to primary tasks solved within radar signal processing. In this paper, Taylor series based tracking algorithm for through wall tracking of moving target by UWB radar system is introduced. It is derived from Taylor series method, which uses actual time of arrival measurements for target localization. In contrast to Taylor series method, Taylor series based tracking algorithm for target positioning exploits not only actual time of arrival measurements, but also estimated target position obtained in the previous time instant by Taylor series based tracking algorithm. In order to improve the tracking ability of the proposed algorithm, a suitable weighting of input data of the algorithm is applied. The performance of Taylor series based tracking algorithm will be compared with performance of direct calculation method and linear Kalman filtering. For that purpose, processing of signals obtained by M-Sequence UWB radar system at through wall measurement has been used. The obtained results will show very clearly that the Taylor series based tracking method can provide better estimate of the target trajectory than the other tested localization and tracking algorithms.

**Keywords**—Taylor series based tracking algorithm, target tracking, UWB radar system, TOA.

## I. INTRODUCTION

A radar system can be inter alia used for tracking of the target. In general, radar systems use modulated waveforms and antennas to transmit electromagnetic energy into a space to search the target. The target within a search will reflect portions of this energy back to the radar system. These echoes are then processed by the radar receivers to extract target information such as velocity, location and other target identifying characteristics.

Ultra wideband (UWB) radar systems operating in a lower GHz-range base-band (up to 5 GHz) are characteristic with good penetration through various obstacles, e.g. through most common building materials including concrete, lite concrete, brick, wood, plastic, tile, and glass, as well as through ground or snow. Such radar systems can be therefore very helpful for applications including security (localization authorized persons in high-security areas or unauthorized persons), family communications (supervision of children), search and rescue works (detecting people buried by avalanche or earthquake), etc. [1].

Tracking of the moving target, i.e. determining target coordinates as the continuous function of the time, by using radar system is the complex process that includes three phases [2]: target detection, the distance estimation among transmitting antenna (Tx), target (T) and receiving antenna (Rx) and tracking itself. The decision, if the target is or is not present, is the output of the detection phase. If the object is detected, the distance among transmitting antenna, target and receiving (Tx-T-Rx) can be estimated. For that purpose,

the time of arrival (TOA) among Tx-T-Rx can be used with advantage for UWB radar systems. The last phase of the radar processing is tracking, i.e. determining target coordinates in defined coordinate system based on estimated distances among Tx-T-Rx in every observation time instant.

There are two basic approaches how to track moving target. The first one consists of localization, i.e. determining target coordinates in every observation time instant separately from TOA measurements and tracking itself. The target locations estimated in consecutive time instants create object trajectory. The object trajectory estimated by the localization methods can be further processed by tracking algorithms. They provide a new estimation of the object localization based on actual and previous positions of the target. Usually, the target tracking will result in the target trajectory error estimation decreasing including trajectory smoothing. There are several iterative and non-iterative localization methods, overviews of which can be found e.g. in [3], [4], [5]. On the other hand, various kinds of Kalman filters are the most frequently used for tracking in radar technology [6].

The second approach how to determining moving target coordinates is tracking making use of actual TOA measurements and foregoing target positions. In this paper it will be describe a new tracking algorithms based on Taylor series. It uses the actual TOA measurements and distances among Tx, target in previous time instant and Rx. Its facilities have been verified on real data. The results have been shown the better and smoother estimate of the target trajectory.

The structure of the paper is as follows. In the Section II, the system model will be described. The detailed description of the Taylor series based tracking algorithms will be described in Section III. Then, in Section IV Taylor series based tracking algorithm ability will be illustrated and compared with that of direct calculation method and linear Kalman filtering. For that purpose, the results of through wall tracking of moving person by using M-Sequence UWB radar at two different scenarios will be used. Finally, some extensions of Taylor series based tracking algorithm will be suggested in Section V.

## II. SYSTEM MODEL

Let us consider the UWB radar system with one Tx and two receiving antennas  $Rx_i$ ,  $i = 1, 2$  located in the known positions with coordinates  $Tx = (x_t, y_t)$ ,  $Rx_i = (x_i, y_i)$ ,  $i=1,2$ . All antennas have known positions in 2D. Let  $TOA_i(t)$ ,  $i = 1, 2$  be the measured time of arrival of the electromagnetic wave transmitted from transmitting antenna Tx, reflected by the target and received by the  $i$ th Rx in the time instant  $t$ . The goal is to determine the unknown coordinates  $(x(t), y(t))$  of the target in 2D for every observed time instant  $t$ .

The distances among Tx-T-Rx<sub>*i*</sub> for  $i = 1, 2$  can be computed from measured  $TOA_i(t)$  in time instant  $t$  as

$$d_i(t) = c \cdot TOA_i(t), \quad i = 1, 2 \quad (1)$$

where  $c$  is the electromagnetic wave propagation velocity. In our consideration,  $c$  is set to the electromagnetic wave propagation velocity in air, i.e.  $c = 3 \cdot 10^8$  ms<sup>-1</sup>. In our paper, the so-called wall effect (different velocities of electromagnetic wave propagation through wall and in air) is not taken into consideration.

In the real scenarios, the distances  $d_i(t)$ ,  $i = 1, 2$  are estimated with error represented by the additive noise component  $e_i(t)$ . Then, the estimated distances  $d_i(t)$  can be modelled as

$$d_i(t) = r_i(t) + e_i(t) = \sqrt{(x(t) - x_i)^2 + (y(t) - y_i)^2} + \sqrt{(x(t) - x_t)^2 + (y(t) - y_t)^2} + e_i(t), \quad i = 1, 2 \quad (2)$$

where  $e_i(t)$  are noises and  $r_i(t)$ ,  $i = 1, 2$  are true distances among Tx-T-Rx<sub>*i*</sub>.

The target localization by using Taylor series method (TSM) was originally proposed in [7]. Then in [8], the original TSM was modified for target localization based on TOA measurements among Tx-T-Rx. However, this method gives for such radar systems the same results as simple localization based on direct calculation method [8], [9]. Direct calculation method for such defined radar system computes intersection of ellipses. The reason is small number of TOA measurements, i.e. small number of equations (2). The solution is to add new equations or additional TOA measurements.

### III. TAYLOR SERIES BASED TRACKING ALGORITHM

As it was outlined above, the Taylor Series Based Tracking Algorithm (TST) will be derived from TSM. TSM uses only actual TOA measurements for target localization. TST in contrast to TSM, uses for target localization not only actual TOA measurements, but also estimated target position obtained by TST in the previous time instants.

The TST for calculation of unknown coordinates  $(x(t), y(t))$  of the target in the time instant  $t$  consists of these steps:

- 1) The set of estimated distances (2) among Tx-T-Rx<sub>*i*</sub> for  $i = 1, 2$  in the actual time instant  $t$  is extended about two new distances

$$d_i(t) = \sqrt{(\hat{x}(t-1) - x_i)^2 + (\hat{y}(t-1) - y_i)^2} + \sqrt{(\hat{x}(t-1) - x_t)^2 + (\hat{y}(t-1) - y_t)^2} \quad (3)$$

for  $i = 3, 4$  where  $(\hat{x}(t-1), \hat{y}(t-1))$  are estimated coordinates of target in the previous time instant  $t-1$ . All distances are then modelled as

$$d_i(t) = \sqrt{(x(t) - x_i)^2 + (y(t) - y_i)^2} + \sqrt{(x(t) - x_t)^2 + (y(t) - y_t)^2} + e_i(t), \quad i = 1, 2, 3, 4. \quad (4)$$

- 2) The non-linear equations (4) are linearized by their expanding in Taylor series [10] around a point (which is subsequently estimated within iteration process) and keeping only terms below second order.

Let us set the initial estimate of the target coordinates as  $(x_v, y_v) = (\hat{x}(t-1), \hat{y}(t-1))$  and define new functions

$$f_i(x, y) = \sqrt{(x - x_i)^2 + (y - y_i)^2} + \sqrt{(x - x_t)^2 + (y - y_t)^2}, \quad i = 1, 2, 3, 4. \quad (5)$$

Then, equations (5) can be rewritten as

$$f_i(x, y) = d_i(t) - e_i(t), \quad i = 1, 2, 3, 4 \quad (6)$$

where  $e_i(t)$  is distance estimation error.

If  $x_v$  and  $y_v$  are initial estimate of the target coordinates, then

$$x = x_v + \delta_x, \quad y = y_v + \delta_y \quad (7)$$

where  $\delta_x$  and  $\delta_y$  are the positions errors to be determined. Expanding  $f_i$  in Taylor series and retaining the first two terms produces

$$f_{iv} + a_{i1}\delta_x + a_{i2}\delta_y \approx d_i(t) - e_i(t), \quad i = 1, 2, 3, 4 \quad (8)$$

where

$$\begin{aligned} f_{iv} &= f_i(x_v, y_v), \\ a_{i1} &= \frac{\partial f_i}{\partial x} \Big|_{x_v, y_v} = \frac{x_v - x_i}{r_{iv}} + \frac{x_v - x_t}{r_{tv}}, \\ a_{i2} &= \frac{\partial f_i}{\partial y} \Big|_{x_v, y_v} = \frac{y_v - y_i}{r_{iv}} + \frac{y_v - y_t}{r_{tv}}, \\ r_{iv} &= \sqrt{(x_v - x_i)^2 + (y_v - y_i)^2}, \\ r_{tv} &= \sqrt{(x_v - x_t)^2 + (y_v - y_t)^2}. \end{aligned}$$

Equations (8) can be rewritten in matrix form as

$$\mathbf{A}\delta = \mathbf{D} + \mathbf{e} \quad (9)$$

where

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \\ a_{41} & a_{42} \end{bmatrix}, \quad \delta = \begin{bmatrix} \delta_x \\ \delta_y \end{bmatrix},$$

$$\mathbf{D} = \begin{bmatrix} d_1(t) - f_{1v} \\ d_2(t) - f_{2v} \\ d_3(t) - f_{3v} \\ d_4(t) - f_{4v} \end{bmatrix}.$$

- 3) The set of linear equations written by matrix form (9) is solved by weighted least-squares method [10] to produce a new approximate position of the target. The iteration process continues until a pre-defined criterion is satisfied. From (9), the weighted least-squares solution of  $\delta$  with weighting matrix  $\mathbf{W}$  is

$$\delta = [\mathbf{A}^T \mathbf{W} \mathbf{A}]^{-1} \mathbf{A}^T \mathbf{W} \mathbf{D}. \quad (10)$$

For initial target position  $(x_v, y_v)$  the update target position is estimated according to  $x_v = x_v + \delta_x$ ,  $y_v = y_v + \delta_y$  where  $\delta$  is computed by (10). The iteration process is repeated until  $\delta$  is sufficiently small.

Generally, TSM and TST require good initial target guess. If the initial guess is not close to the true solution, divergence may occur. However, the divergence behavior is easily detected. One possible method to detect divergence is to check if  $\|\delta\|$  from the current iteration is larger than that  $\|\delta\|$  in the previous iteration. If this is the case, the process is not converging and we should try a new initial guess.

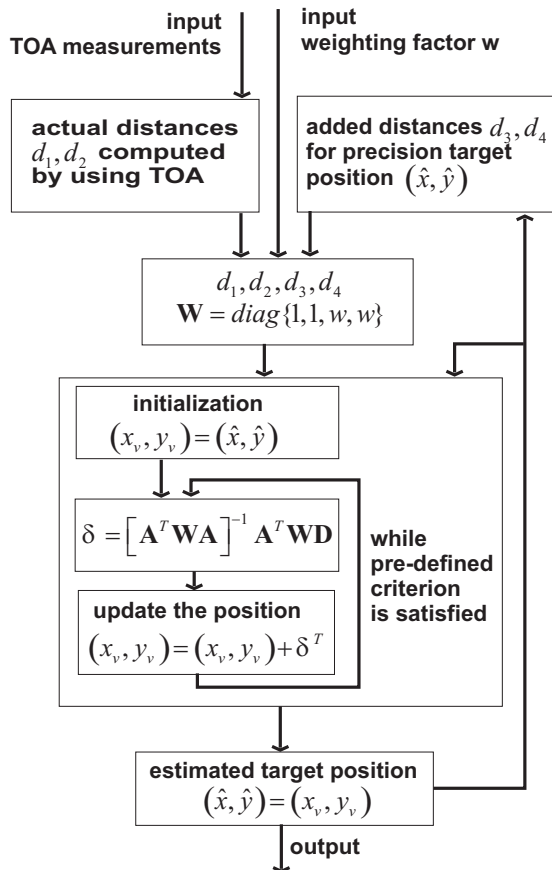


Fig. 1: Computation flow of TST.

Generally, the particular weights of weighting matrix  $\mathbf{W}$  can be used to characterize the reliability of TOA measurements. According to this idea, the more accurate measurements are placed with a larger weight to stress the importance of the more reliable observations.

In TST is the weighting matrix  $\mathbf{W}$  from equation (10) set to

$$\mathbf{W} = \text{diag}\{1, 1, w, w\} \quad (11)$$

where  $w \in \mathbb{R}$  is weighting factor. That matrix weights the actual estimated distances with new added distances computed from precision target position. If  $w < 1$ , then new estimated distances are used in TST with the larger reliabilities than the added distances. On the contrary, if  $w > 1$ , then new added distances are used in TST with the lesser reliabilities than actual estimated distances. If  $w = 1$ , the reliabilities are same.

Bath weighting manner can be connected.

- 4) Then, the estimated target position in actual time instant is  $(\hat{x}(t), \hat{y}(t)) = (x_v, y_v)$ .

The computation flow of TST is shown in the Fig. 1.

#### IV. PERFORMANCE OF TST

The radar data analyzed in this paper were acquired by M-sequence UWB radar [11] with 4.5 GHz system clock and measurement speed approximately 10 impulse responses per second equipped with one Tx and two Rxs. More details concerning radar system devices can be found in [12].

To get TOA estimation row radar data have been processed by the procedures described in [13], i.e. raw radar data pre-processing, background subtraction, detection and trace

estimation. The data after trace estimation in every time instant represent TOA estimation.

In order to illustrate TST performance, TST was applied for processing of radar signal obtained processed at two scenarios where the observation target was a person.

In the first scenario the moving person was walking from the position P(1), through the positions P(2), P(3), and P(4) and back to the position P(1). The measurement was made behind lite concrete wall with thickness 18 cm. The distance between adjacent Rxs was set to 76 cm. The Tx was in the centre of the TxS. All antennas were placed 78 cm above the floor. Other distances are schematically depicted in the Fig. 2. The purpose of this scenario is to compare the performance of different approaches of target positioning, i.e. direct calculation method, linear Kalman filter and TST.

In the second scenario, one person was walking from the position P(1), through the positions P(2), P(3), and P(4) and back to the position P(1). The measurement was made behind the wall covered by tile with thickness 24 cm. The distance between adjacent Rxs was set to 260 cm. The Tx was in the centre of the Rxs. All antennas were placed 120 cm above the floor. Other distances are schematically depicted in the Fig. 4. By using the data, obtained in this scenario, we will illustrate TST performance depending on selection of weighting factor. For this scenario, we will compare TST performance for different weights.

##### A. Comparison of TST with the direct calculation method and linear Kalman filtering

By using estimated distances computed from TOA by (1) the position of the moving target for every observation time instant has been determined. For that purpose, the localization by using direct calculation method [8] was used. The target localization results are depicted by thin curve in the Fig. 3.

In order to track moving target, linear Kalman filter [6] and TST have been used. The results of tracking are given in the Fig. 3, where the trajectory estimation by linear Kalman filter and TST are represented by dotted and thick curve, respectively. For this scenario, the weighting factor from (11) was set to  $w = 5$ .

It can be seen, that the positions of moving person computed by TST are more precise than that of positions computed by simple localization by using direct calculation method or by linear Kalman filter.

##### B. TST performance depending on selection of weighting factor

In the second scenario, the comparison of TST depending on selection of weighting factor has been made. The results are depicted in the Fig. 5 for the weighting factors  $w = 1, 5, 10$ , respectively. The depicted stars represent the positions of the person computed by the direct calculation method for every observation time instant separately.

It can be observed, that the TST for higher weights can provide the more smoothed line as the TST for lower weights. On the contrary, if the person changed the direction of the motion, TST for higher weighting factor gives the less precise results. This effect can be identified especially at the points P(1), P(2), P(3), P(4) of target trajectory.

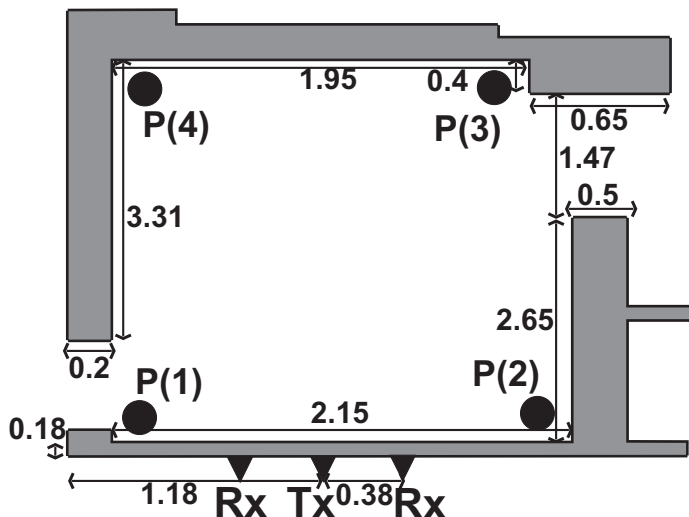


Fig. 2: First measurement scenario.

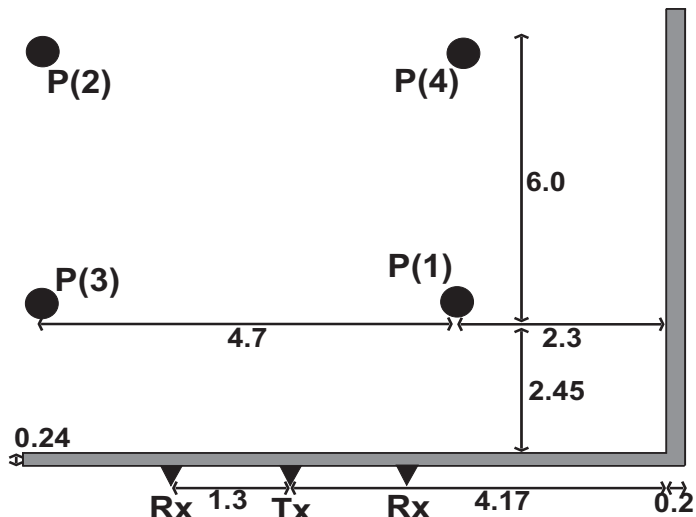


Fig. 4: Second measurement scenario.

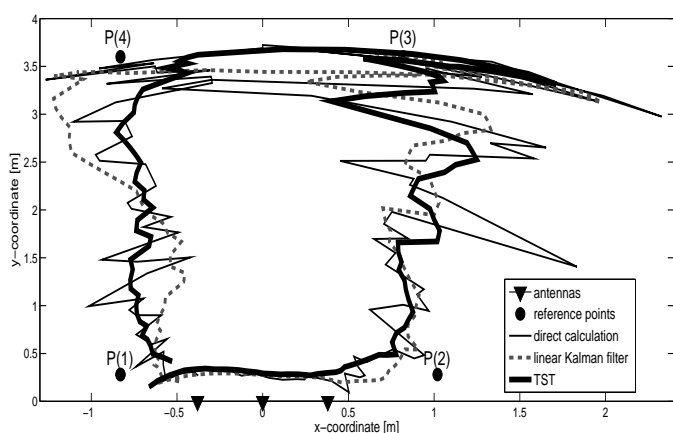


Fig. 3: Scenario 1: Target trace estimated by using direct calculation method, linear Kalman filter and Taylor Series Based Tracking Algorithm.

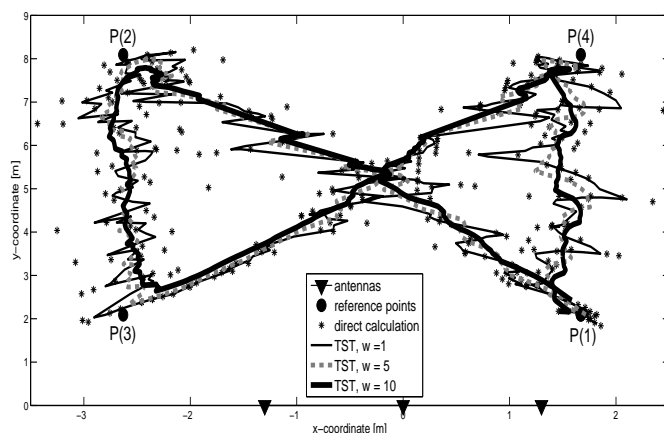


Fig. 5: Scenario 2: Target trace estimated by using Taylor series based tracking algorithm depending on weighting factor  $w$ .

V. CONCLUSION

In this paper the Taylor series based tracking algorithm was described. Our next research concerning this algorithm will be intent on development of TST modification for target tracking by UWB sensor network. It is expected that the application of modified version TST in combination with UWB sensor network can improve the estimation of the target trajectory in a significant way.

ACKNOWLEDGMENT

This work was supported by the Slovak Research and Development Agency under the contract No. LPP-0287-06 and by European Commission under the contract COOP-CT-2006-032744.

REFERENCES

[1] I. Oppermann, M. Hamalainen, and J. Iinatti, *UWB Theory and Applications*. John Wiley & Sons, Ltd, England, November 2004.  
 [2] J. Rovňáková, M. Švecová, D. Kocur, T. T. Nguyen, and J. Sachs, "Signal Processing for Through Wall Moving Target Tracking by M-sequence UWB Radar," *The 18th International Conference Radioelektronika, Prague, Czech Republic*, pp. 65–68, April 2008.  
 [3] M. Švecová, "Node Localization in UWB Wireless Sensor Networks," Thesis to the dissertation examination, Technical University of Košice, Department of Electronics and Multimedia Communications, Slovak Republic, December 2008.

[4] K. Yu, J.-p. Montillet, A. Rabbachin, P. Cheong, and I. Oppermann, "UWB Location and Tracking for Wireless Embedded Networks," *Signal Processing*, vol. 86, no. 9, pp. 2153–2171, September 2006.  
 [5] M. Švecová, "Node Localization Methods in UWB Wireless Sensor Networks: A Review," *8th Scientific Conference of Young Researchers FEI TU of Košice, SCYR 2008*, May 2008.  
 [6] M. S. Grewal and A. P. Andrews, *Kalman Filtering: Theory and Practice Using MATLAB*, 3rd ed. Wiley-IEEE Press, September 2008.  
 [7] W. H. Foy, "Position-Location Solutions by Taylor-Series Estimation," *IEEE Transaction on Aerospace and Electronic Systems*, vol. AES-12, no. 2, pp. 187–194, March 1976.  
 [8] M. Švecová, D. Kocur, and R. Zetik, "Object Localization Using Round Trip Propagation Time Measurements," *The 18th International Conference Radioelektronika, Prague, Czech Republic*, pp. 41–44, April 2008.  
 [9] M. Aftanas, J. Rovňáková, M. Rišková, D. Kocur, and M. Drutarovský, "An Analysis of 2D Target Positioning Accuracy for M-sequence UWB Radar System under Ideal Conditions," *The 18th International Conference Radioelektronika, Brno, Czech Republic*, pp. 189–194, April 2007.  
 [10] E. K. P. Chong and S. H. Zak, *An Introduction to Optimization*, 3rd ed. Wiley-Interscience, February 2008.  
 [11] M. Riskova, J. Rovnakova, and M. Aftanas, "M-Sequence UWB Radar Architecture for Throughwall Detection and Localisation," *VII. National Science Conference at FEI, Technical University in Košice, June 2007*.  
 [12] D. Daniels, "M-sequence radar," in *Ground Penetrating Radar*. London, United Kingdom: The Institution of Electrical Engineers, 2004.  
 [13] J. Rovňáková, "UWB Signal Processing for Moving Target Detection," Thesis to the dissertation examination, Technical University of Košice, Department of Electronics and Multimedia Communications, Slovak Republic, December 2008.

# Methods of Realization of Innerlayer Cavities in Low Temperature Cofired Ceramics

<sup>1</sup>Igor VEHEC, <sup>2</sup>Pavol CABÚK

Dept. of Technologies in Electronics, FEI TU of Košice, Slovak Republic

<sup>1</sup>i.vehec@tuke.sk, <sup>2</sup>pavol.cabuk@tuke.sk

**Abstract**—Low Temperature Cofired Ceramics (LTCC) offers many features for fabrication of compact 3D structures. This paper is focused on realization of innerlayer cavities. There were chosen four methods for cavities realization: isostatic lamination, isostatic lamination with stainless steel plate on the top of samples, isostatic lamination with carbon powder and low temperature and pressure lamination. Advantages and disadvantages of these methods were compared at the conclusion.

**Keywords**—LTCC (Low Temperature Cofired Ceramics), carbon powder, innerlayer cavities.

## I. INTRODUCTION

Low Temperature Cofired Ceramics (LTCC) is progressive evolving ceramic material in the field of electrotechnical engineering. Its advantages are primarily low temperature of firing (under 1000°C) and flexibility in green (unfired) state, that facilitate realize a planar and a shaped multi-layers structures.

The possibility of compact structures fabrication based on LTCC ceramic is one of their major domains. In the present, flexibility of ceramic in green state is just the most used for different application in fields of sensors, MEMS (Micro-Electro Mechanical Systems) and MST (Micro-System Technology).

The LTCC ceramic in green state allows make manifold shaped cavities, channels, embedded structures and other ones. The cavities can be situated on the top of structure (toplayer cavities) or can be embedded in the structure (innerlayer cavities).

The toplayer cavities are more easily to make as innerlayer cavities. Their applications are for sensor carriers, area for chip assembly, etc.

The innerlayer cavities are used e.g. for realization channels or fine capillaries for cooling or micro-fluidic applications, etc. The innerlayer cavity basically consists from structure with toplayer cavity, whereon stacked additional layers. These layers can overlap the top cavity fully or partially.

## II. THEORETICAL BACKGROUND

Sagging of layers (over and under innerlayer cavity) during processing of LTCC ceramic with innerlayer cavities due to

deformation during lamination or due to thermal tension at glass transmission temperature during sintering is the common problem.

However, major technological step which determine quality of innerlayer cavities is lamination at increased temperature and pressure (thermo-compress lamination). The lamination can be uniaxial or isostatic. The isostatic lamination is more suitable method of lamination, because we can obtain high homogeneity of material density due to uniform distribution of pressure from all sides. This proves to more uniform shrinkage after firing and better hermeticity [1], [2].

Quality of the innerlayer cavities depends on magnitude of applied pressure, temperature, respectively on simultaneous influence of pressure and temperature during thermo-compress lamination. The material undergoes viscoelastic deformation which affects its shrinkage (densification) and increasing of material density. When the cavity is located inside stacked structure, distribution of stressing is not uniform. This leads to sagging of top and bottom cavity layers and convexity of cavity sidewalls. The material under pressure flows in x-y direction, and excess material flows to unsupported areas on the top and bottom of cavity. The material flowing to unsupported areas is cause of the material sagging [2].

There was developed several methods for elimination of top and bottom layers sagging problem due to the flowing materials in innerlayer cavity. Among these methods belongs: lamination with using lamination inserts, multi-stage lamination [3], using sacrificial materials removed by etching and using sacrificial materials removed by burn-out [4].

## III. EXPERIMENTAL PART

All experiments were realized with LTCC ceramic Green Tape™ 951AX with thickness 254 μm. The samples consist of 6 layers of LTCC ceramic with dimension of 14 × 24 mm. The cavities were formed by two inner layers with holes. One sample contains 4 cavities with length of 10 mm and width of 2.0, 3.0, 4.0 and 5.0 mm (Fig. 1a). Different width of the cavities facilitated analyzing its influence on rate of sagging of the covering (resp. underlying) layers. The two underlying layers are created without any holes (Fig. 1b).



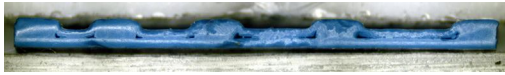

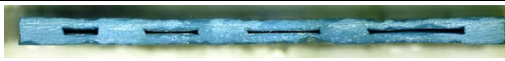

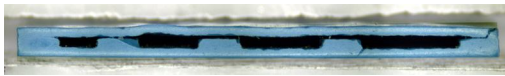



preparation.

### B. Results of experiments and discussion

Suitability and unsuitability of chosen methods of innerlayer cavities realization were considered on the basis of top layers (resp. bottom layers) planarity. Microscopic analyzes of samples cross-section are demonstrated in the Tab. I.

TABLE I  
EXPERIMENTAL RESULTS OF INNER CAVITY REALIZATION

Sample	Result
Isostatic lamination	
A	
B <sup>1</sup>	
Isostatic lamination with stainless steel plate on the top of samples	
C	
Isostatic lamination with carbon powder	
D	
E <sup>2</sup>	
Low temperature and pressure lamination	
F	

<sup>1</sup>Width of cavities: 0.5 mm, 1.0 mm and 2.0 mm (from left to right), numbers of top layer: 4.

<sup>2</sup>Exceeded deposition of carbon powder.

(a) Isostatic lamination (Tab. I – samples A and B). During isostatic lamination material flows and this is case of top layers sagging. Bottom layers are planar, because they were supported by rigid support. The top layers deformation was observed over all chosen widths of samples (2.0, 3.0, 4.0 and 5.0 mm) on sample A. Experiment was repeated for sample with cavities widths of 0.5, 1.0 and 2.0 mm and 4 top layers (sample B – Tab. I). In this case, only cavity with width of 0.5 mm is kept planarity of top layers. As we can see, the sagging depends on the cavity width as well as on the numbers of layers over cavity. This method is suitable only for narrow cavities with enough numbers of top layers.

(b) Isostatic lamination with stainless steel plate on the top of samples (Tab. I – sample C). The sagging of top layers occurs in spite of the steel plate placed on the top surface of sample. Moreover, the sagging of bottom layers was occurred (the sample was placed on rigid support during lamination). In comparison with previously methods, this method can give more planarity and higher clearness of cavity.

(c) Isostatic lamination with carbon powder in the cavities (Tab. I – samples D and E). The carbon powder burn-out for the samples Type II (samples with circular cavities in top layers) as well as for samples Type I (samples without any cavities in top layers). In the cavities remains only small

amount of impurities associated with purity of used carbon powder. Experiment with burn-out of used carbon powder in oxygen atmosphere showed that its burn-out goes at temperatures between 500°C and 600°C. Results of experiments mentioned in [7] point to that ceramic keeps its porosity at temperature under ~750°C, so carbon powder can burn-out through those open porosity. Over this temperature, the pores start to close until the ceramic is no more permeable. This occurs at temperature ~785°C.

Using carbon powder is good method for preventing of sagging unsupported layers (Tab. I – sample D). But, the results are dependent on: (1) suitable ratio of carbon powder and organic binder, and (2) suitable method for paste deposition.

(ad 1.) Because there must be deposited huge amounts of paste, for big ratio of organic vehicle there need prolong time of drying. This can affect on lamination in consequence of over-drying of ceramics. Insufficient drying on the other side can damage ceramics due to organic vehicle absorption of the tape.

(ad 2.) The second problem is suitable method for deposition of paste. The paste must be precisely applied into the cavity so that paste amount does not too much or deficient. The influence of excessive amount of carbon paste is showed on the picture of sample E in Tab. I. The paste got out of cavity during lamination and caused that upper and bottom part of sample were not laminated. There was observed continuous interconnection of cavities after firing. On the other side, insufficient amount of paste can lead to sagging of top layers during lamination.

The next problem of carbon powder deposition is planarity of deposited paste, because ceramics in green state copy rough surface during lamination. This copying is more visible for lesser numbers of top layers.

TABLE II  
ADVANTAGES AND DISADVANTAGES OF SOME METHODS OF INNER CAVITIES REALIZATION

Used technology	Advantage (+) / Disadvantage (-)
Isost. lamination	+ simply methods without using additional materials, - methods is suitable only for cavities with small width or/and sufficient numbers of top layers,
Isost. lam. with stainless steel plate on the top of samples	+ better planarity and clearness of cavities, - sagging of layers on top as well as on bottom of cavity,
Isost. lam. with carbon paste	+ suitable methods for making cavities and channels, - sensitive on the used paste and technology of application paste,
Low temperature and pressure lam.	+ suitable methods for making cavities and channels, - possibility of top layers sagging during firing of samples with wider cavities,

(d) Low temperature and pressure lamination with using organic fluid (Tab I – sample F). The last method offers the best planarity top as when as bottom layers. There was not observed sagging of top layers during firing for chosen width

of cavities as well. But for this method is necessary another analyses for mechanical properties as well as hermeticity of structures.

Advantages and disadvantages for chosen methods of inner cavity realization are summarized in Tab. II.

#### IV. CONCLUSION

There were compared four methods of realization of innerlayers cavities. The choice of suitable methods depends on cavities width, numbers of top layers and requirements on planarity of top (resp. bottom) surface of samples. For small cavities is there sufficient multi-stage isostatic lamination or isostatic lamination with stainless steel plate on the top of samples. For larger width of cavities it is necessary to choice the more sophisticated methods: lamination with sacrificial layers (e.g. carbon powder) or low temperature and pressure lamination. (VEGA No. 1/0108/09)

#### REFERENCES

- [1] DuPont Green Tape™ Design and Layout Guideline, www.dupont.com (2004)
- [2] M.A. Zampino, *Embedded Heat Pipes in Cofired Ceramic Substrates for Enhanced Thermal Management of Electronics* (dissertation), Florida International University Miami, Florida 2001.
- [3] R. Bauer, M. Luniak, L. Rebenklau, K.J. Wolter, W. Sauer, *Realization of LTCC-Multilayer with Special Cavity Applications*. 30<sup>th</sup> International Symposium on Microelectronics, ISHM 97, Philadelphia, October 1997.
- [4] M.R. Gongora-Rubio, P. Espinoza-Vallejos, L. Sola-Laguna, J.J. Santiago-Avilés, *Overview of low temperature co-fired ceramics tape technology for meso-system technology (MsST)*. Sensors and Actuators A 89, Amsterdam 2001, pp. 222 – 241
- [5] Z.M. da Rocha, N.I. Garcia, N.A. de Oliveira, J. do Rosário Matos, M.R. Gongora-Rubio: *Low Temperature and Pressure Lamination of LTCC Tapes for Meso-systems*, IMAPS Conference and Exhibition on Ceramic Interconnect Technology, Denver 2004
- [6] M.A. Piwonski, A. Roosen, *Low Pressure Lamination of Ceramic Green Tapes by Gluing at Room Temperature*. Journal of the European Ceramic Society 19 (1999), Elsevier, pp. 263-270
- [7] H. Birol, T. Maeder, P. Ryser, *Processing of Graphite-Based Sacrificial Layer for Microfabrication of Low Temperature Co-fired Ceramics (LTCC)*. Sensors and Actuators A: Physical, vol. 130-131 (2006), Elsevier, p. 560-567, 2006

# Motor Speed Regulation via Internet and Artificial Neural Network

Tibor VINCE

<sup>1</sup>Dept. of Theoretical Electrical Engineering and Electrical Measurement, FEI TU of Košice, Slovak Republic

tibor.vince@tuke.sk

**Abstract**—The article presents regulation possibilities of DC motor via Internet and possible improvements of such control using Artificial Neural Network. The paper also handles the advantages and disadvantages of Internet as a control and communication bus in different level of the information hierarchy. The article handles also with artificial neural network utilization in such a regulation. The contribution contains also such a remote regulation proposal and presents also remote regulation response measured data.

**Keywords**—Remote control, Internet, Artificial Neural Network, DC Motor, Information architecture

## I. INTRODUCTION

There is huge effort to integrate different cooperating systems in one complex system. Basic problem is communication between these different modules of the system, especially when the modules are located in different locations. According to communication requirements, appropriate communication way has to be chosen.

Continual evolution of the Internet enables higher and higher communication requirements to be fulfilled. The Internet begins to play a very important role in industrial processes manipulation, not only in information retrieving. With the progress of the Internet it is possible to control and regulate from anywhere around the world at any time. New concept of controlling, which has been paid much attention in these years is distance remote via Internet, or other words, Internet-based control.

Such a type of control bus allows remote monitoring or regulation of whole plants or single devices over the Internet. The design process for the Internet-based control systems includes requirement specification, architecture design, control algorithm, interface design and possibly safety analysis. Due to the low price and robustness resulting from its wide acceptance and deployment, Ethernet has become an attractive candidate for real-time control networks.

In some case in such remote regulation is necessary to regulate electric motor. The goal of the article is to explore now days possibilities for Internet based real-time regulation, eventual trends, review of advantages and disadvantages of distance remote via Internet in different level of information hierarchy and possible solutions. The article presents regulation of DC motor as an example such a real-time regulation and discusses possibilities utilization Artificial Neural Network (as part of Artificial Intelligence). On the end

it presents some measured data from such a remote DC motor regulation via Internet.

## II. INFORMATION ARCHITECTURE

As mentioned before, there is effort to integrate different subsystems in one complex system. Integration of information and control across the entire plant site becomes more and more significant. In the manufacturing industries this is often referred to as "Computer Integrated Manufacturing" (CIM). There is increasing use of microprocessor-based plant level devices such as programmable controllers, distributed digital control systems, smart analyzers etc. Most of these devices have "RS232" connectors, which enable connection to computers. If we began to hook all these RS232 ports together, there would soon be an unmanageable mess of wiring, custom software and little or no communication. This problem solution results in integration these devices into a meaningful "Information Architecture". This Information Architecture can be separated into 4 levels with the sensor/actuator level as shown in Fig. 1, which are distinguished from each other by "4Rs" principle criteria: [1]

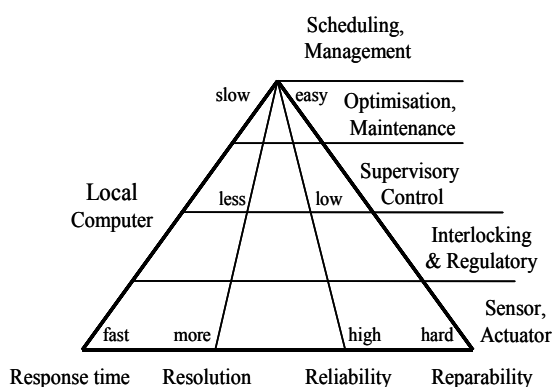


Fig 1. Information Architecture

The 4Rs criteria are: Response time, Resolution, Reliability and Reparability.

**Response time:** as one moves higher in the information architecture, the time delay, which can be tolerated in receiving the data, increases. Conversely, information used at the management & scheduling level can be several days old without impacting its usefulness.

**Resolution:** an Abstraction level for data varies among all

the levels in the architecture. The higher the level is, the more abstract the data is.

**Reliability:** Just as communication response time must decrease as one descends through the levels of the information architecture, the required level of reliability increases. For instance, host computers at the management & scheduling level can safely be shut down for hours or even days, with relatively minor consequences. If the network, which connects controllers at the supervisory control level and/or the regulatory control level, fails for a few minutes, a plant shutdown may be necessary.

**Reparability:** The reparability considers the ease with which control and computing devices can be maintained.

Local computer on supervisory control level is able to communicate with higher levels of information architecture via Internet, but there is also possibility to use the Internet also in lower levels of the Information architecture. The Internet can be linked with the local computer system at any level in the information architecture, or even at the sensor/actuator level. These links result in a range of 4Rs (response time, resolution, reliability, and reparability). For example, if a fast response time is required a link to the control loop level should be made. If only abstracted information is needed the Internet should be linked with a higher level in the information architecture such as the management level or the optimization level.

### III. ARTIFICIAL NEURAL NETWORK UTILIZATION

Since the early 1990's, there has been a growing interest in using artificial neural networks for control of nonlinear systems. Numerous applications have demonstrated that neural networks are indeed powerful tools for the design of controllers for complex nonlinear systems. Among different kinds of neural networks, the most widely used ones are multilayer neural networks and recurrent networks. For the Internet-based Control is very important kind of auto adaptation.

By solving tasks in field of electric drives we meet following basic problems: system simulations, identification of system parameters, system state quantities monitoring, drive regulations and malfunction diagnostic. There is possible successful utilization of neural network in all these fields of problem. The most important neural network attributes in this field are: various nonlinear functions approximation, parameters settings based on experimental or learning data, data processing and robustness.

Two different models are used in identification models creation: mathematics and physics analysis and experimental identification. In case of complex subject both methods are required. Neural network can be used as direct neural model connected parallelly or serial-parallelly in learning state. Neural network can be used also as inverse identification model in dynamic system. Today's computer science performance allows to replace classical methods of parameters estimation by automatic identification. Main advantages are complex test signals generation possibilities, sophisticated identification algorithms, on-line identification possibilities etc.

The condition of the Internet is a very varying parameter and the control system controlling via Internet has to compensate

the variation. One of the solutions is to employ the neural network. It is possible to teach neural network behavior for different conditions of networks. Advantage of this solution is that neural network is a more universal tool and the condition of the Internet can be used as one of many parameters related to controlling and regulation.

### IV. REMOTE REGULATION PROPOSAL

Adequate control software, appropriate computers on client and server site and Internet with satisfactory connection speed are necessary for successful remote regulation of DC motor via Internet. Adequate control software, computer and connection speed depends on concrete motor parameters. In general, regulation of DC may be considered as a real-time regulation problem and time intervals are in tens of milliseconds. The time intervals may vary significantly from every regulation system. For regulation system via Internet is very important if the regulation loop time interval must be under one millisecond, in milliseconds or may be over hundreds of milliseconds and more. In the architecture design, a remote regulation of electric motor via Internet generally includes three major parts: client, server and regulated electric motor. The general remote regulation system architecture is shown in Figure 2. The client part is the interface for the operations.

It includes computers, control software with user interface for operators or superior system. Client computer receives state information of electric motor, connection state and other information related to the motor regulation via Internet. Received information will be processed and evaluated in remote computer.

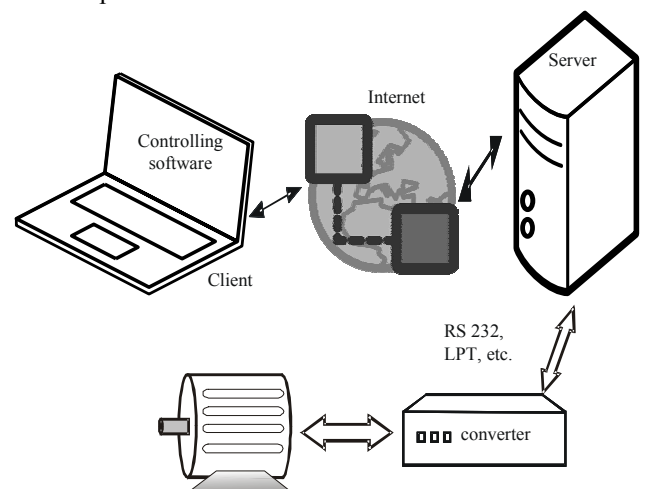


Fig 2. Remote system architecture

The server part contains a server computer, which is connected to the converter. The server contains all required drivers and devices for communication with the converter.

Communication of server with converter could be based on several ways (RS232, Profibus, CAN, USB, etc.). Sophisticated converters may be Ethernet enabled and may be connected directly to the Internet. But if the client computer is located in an outside network – not in a LAN network, where the converter is located, the server computer is recommended.

The third part of system architecture is the electric motor

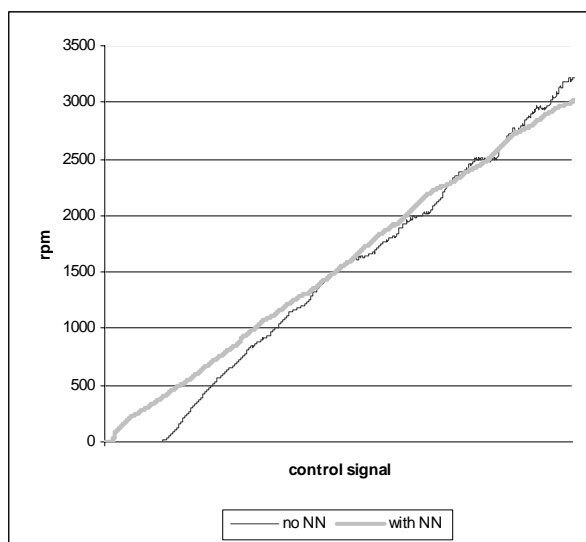
with the converter itself. Common way for distance regulation is, when remote client computer has limited functions – only start/stop of electrical motor, or choose the wished value of speed, torque etc. The regulator itself (for instance PI regulator) is located on server site, or is implemented in converter.

But Internet speed progress open possibility for real-time control from client site, so there is possibility that Internet could be part of the regulation loop. Between client computer and server could be thousands of kilometers, or they could be in the same room. The difference is in the communication delay, but generally the system is the same. The communication service (the bus) can be achieved by wired connection, mostly Ethernet, or wireless – very popular WiFi.

## V. MEASUREMENTS

This chapter presents measured data. The Artificial Neural Network (NN) was used for linearization of motor speed (rpm) according to control signal.

Graph 1 presents system response - rpm on control signal.



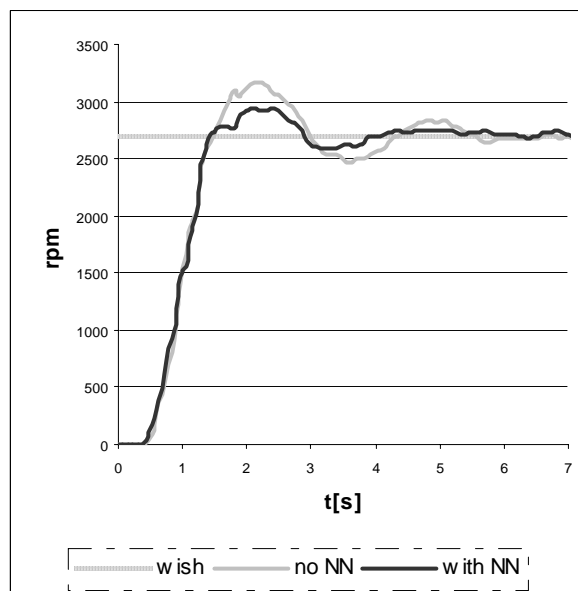
Graph 1. System response

The darker curve presents system response without Neural Network. The lighter curve presents response with Neural Network. As Graph1 shows, response with NN has more advantages: lighter curve shape is more linear, there is smaller zone of insensitive, when system doesn't response on control signal, when control signal is too small. Significant advantage is that NN could take in consider such a constraint as maximum rpm etc. In this case the maximum rpm should be 3000 rpm.

Graph2 shows system response on remote regulation in 15ms delay. System is regulated by PI regulator tuned for local regulation –doesn't take in consider a time delay. As Graph2 shows, darker shape – using NN has lower overregulation and faster reach wished value.

## VI. CONCLUSION

Many control elements have been embedded with Internet-enabled functions. There is possibility that regulation system



Graph 2. Regulation with 15ms delay

could be connected directly to the Internet (without a necessity of a server computer). When compare Ethernet as a bus with other standard types, the most powerful advantage is nearly unlimited size of bus, possible huge distance, open system of the Internet protocols and accessibility of the Internet. As measured data shows, utilization of Artificial Neural Network in remote regulation could be useful contribution and could help to overcome some disadvantages the Ethernet as a bus.

## ACKNOWLEDGMENT

The paper has been prepared by the support of Slovak grant projects VEGA No. 1/4174/07, VEGA No. 1/0660/08, KEGA 3/5227/07, KEGA 3/6386/08 and KEGA 3/6388/08.

## REFERENCES

- [1] [1] Yang, S.H., Tan, L.S., Chen X: Requirements Specification and Architecture Design for Internet-based Control Systems, proceedings of the 26th Annual International Computer Software and Applications Conference (COMPSAC'02), 2002.
- [2] Kweon, S. K, Cho M., Shin K. G.: Soft Real-Time Communication over Ethernet with Adaptive Traffic Smoothing, IEEE Transactions on parallel and distributed systems, VOL. 15, NO. 10, October 2004.
- [3] <http://www.anybus.com/technologies/technologies.shtml>, 10.2.2008
- [4] Kováč, D., Kováčová, I., Molnár, J.: Elektromagnetická Kompatibilita-meranie, ISBN 978-80-553-0151-8
- [5] Tomčík, J., Tomčíková, I.: IT bezpečnosť automatizačných a SCADA systémov (1). In: AT&P Journal, roč. 13, č. 4(2006), s. 50 – 54. ISSN 1336-5010
- [6] Vince, T., Molnár, J., Tomčíková, I.: Remote DC motor speed regulation via Internet. In: OWD 2008: 10<sup>th</sup> International PhD workshop: Wisla, 18-21 October 2008, pp. 293-296. ISBN 83-922242-4-8
- [7] Kováč, D., Kováčová, I., Vince, T.: Elektromagnetická Kompatibilita, ISBN 978-80-553-0150-1.
- [8] Molnár, J., Kováčová, I.: Distance remote measurement of magnetic field, Acta Electrotechnica et Informatica, 2007, No.4, Vol.7, pp. 52-55, ISSN 1335-8243
- [9] Vince, T., Kováčová, I.: Distance control of mechatronic systems via Internet, Acta Electrotechnica et Informatica, 2007, No.3, Vol.7, pp. 63-68, ISSN 1335-8243
- [10] Kováčová, I., Kaňuch, J., Kováč, D.: DC permanent magnet disc motor design with improved EMC, Acta Technica CSAV, Vol. 50, No.3, 2005, pp.291-306, ISSN 0001-7043

## **2<sup>nd</sup> section: Informatics & Telecommunications**

# Knowledge in Software Architectures

Iveta ADAMUŠČÍNOVÁ

Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

iveta.adamuscinova@tuke.sk

**Abstract**—Nowadays, the various approaches to building the software systems that are easily, or even automatically adaptable to permanent changes during their life cycle, present one of the most observed area in scientific research and related trend is, among others, an effort to find and integrate techniques or concepts supporting this direction, even from other engineering disciplines. In this paper, there is presented a principle of the so-called architectural knowledge, an emerging approach within fields of software maintenance and evolution designed to handle some specific problems related to realizations of systems' changes. This paper also includes an analysis of the main aspects dealing with problems related to integration of architectural knowledge into the software system's architecture and briefly outlines some potential solutions.

**Keywords**—architectural knowledge, software maintenance, software evolution, software architecture

## I. INTRODUCTION

Software engineering went through some significant and fundamental changes during the last two decades. According to [1], it could be stated that today its direction is in majority influenced by two main aspects. The first one is concerned with an effort of adjusting to growing software complexity by raising the level of abstraction through modeling (e.g. model-driven development paradigm) and higher-level programming languages. The second aspect is concerned with an effort of adjusting to increasing demand after software systems that are relatively easily, or even automatically adaptable to constantly changing requirements and environment in which they are used, often during their operation.

The fast adaptation to permanent changes of requirements becomes a great competitive advantage which leads to a fact that changes concerning the software systems occupy one of the most important places within software life cycle. However this aspect interposes higher demands on efficiency of processes related to maintenance, evolution and management of software systems. Today, software maintenance presents the most expensive part of the software system life cycle (the surveys indicate that it consumes 60% to 90% of the total life cycle costs) [2]. Program comprehension, impact analysis and regression testing are the most challenging problems of software maintenance [3].

Problems related to efficient realization of changes within the software system during its use, present the considerable motivation for integration of new approaches, specifically integration of several types of knowledge [1].

In this paper, there is presented a principle of an architectural knowledge, one of the approaches designed to overcome problems concerning the demanding processes in

the phases of maintenance and evolution. There are also highlighted some of the major issues related to extraction and integration of this type of knowledge within the software architectures, inhibiting it from its further use.

## II. KNOWLEDGE IN SOFTWARE MAINTENANCE AND EVOLUTION

### A. Principle of Architectural Knowledge

Knowledge presents a key concept within knowledge engineering. It can be defined as following [4]:

*Knowledge presents understanding of a subject area. It includes concepts and facts about that subject area, as well as relations among them and mechanisms for how to combine them to solve problems in that area.*

The specific type of knowledge which represents the current direction towards securing the efficiency of handling the system's changes during the various phases of software life cycle, is the so-called **architectural knowledge**, i.e. knowledge about some specific software system and its environment. This type of knowledge is usually hidden within the artifacts of software system and it is assumed that its explicit representation and following use could significantly influence not only the quality of software system's development but mainly its following stages of management, maintenance and evolution.

Architectural knowledge is defined as following [5]:

*Architectural knowledge presents understanding of specific software system. It is defined as the integrated representation of the software architecture, the architectural design decisions, its application logic and the external context/environment.*

The need for this type of knowledge originates from the well known fact that the basic requirement for effective problems solving during the challenging stages of software life cycle is its detailed **understanding** and that often the best source of knowledge and, in some cases, the only source of knowledge about existing software system is the system itself [6].

### B. Components of Architectural Knowledge

Among the *critical knowledge* related to processes of management, maintenance and evolution of the system, summarily called the architectural knowledge, belong the knowledge about [7], [8]:

- *mutual relations and dependencies between the artifacts of software system*, which help to maintain the software system in the consistent state even after the implementation of required change. Consistency is an inevitable condition for achieving the non-problematic evolution of software systems.
- *mutual relations and dependencies between the elements within an artifact of software system*, which are necessary for analyzing the impacts of processed modifications on other parts of the system.
- *processed modifications within previous versions of the software system*.
- *connections and relations between the software system and its environment* (e.g. other software systems), which express the way of communication and dependencies of these systems and are also necessary for analyzing the impacts of processed internal modifications on external systems.

These kinds of critical knowledge are always very tightly related to the specific software system and they are contained in all its artifacts [5], [6] – in its models, source code, diagrams, documentation, etc. – that are present collectively in system's project database, from which they are accessible in case of any need [8].

The main problem of this approach is that the architectural knowledge, i.e. knowledge about the system, is in principle stored separately of the system itself. This approach leads to significant increase of the complexity of the whole process of accessing, searching, sorting and applying relevant knowledge and consequently also the process of implementing the required modifications with regards on preserving the consistency between the system's artifacts [8], [9].

### III. MAIN PROBLEMS RELATED TO USING ARCHITECTURAL KNOWLEDGE

In previous paragraph, I mentioned one of the major obstacles related to using architectural knowledge within the phases of software life cycle. The potential solution of this problem includes integration of the architectural knowledge directly into the software system as a part of system's architecture. It is supposed, that this approach would enable more flexible modifications of software systems within the phases of management and maintenance in a faster, simpler and mainly more reliable way.

Presented problem consists in general of two main aspects [9]:

1. Identification and extraction of architectural knowledge from software system's artifacts.
2. Integration of architectural knowledge into software system as a part of system's architecture.

#### A. Identification and extraction of architectural knowledge from software system's artifacts

For achieving the effective realization of changes during the maintenance of the system, its profound understanding is crucial. This understanding consists of knowledge about the structure of the system, relations and dependencies between its

elements, knowledge about the ways these elements can be modified and how the changes affect other part of the system.

Nowadays, it is assumed that one of the best sources of knowledge about specific system is system itself, as it presents not only some source of unrelated data, but also provides useful information about how these data can be processed and used following some intended purpose [6]. To further use of this knowledge, it is necessary to (Fig. 1) [9]:

- *identify the critical architectural knowledge*,
- *extract the architectural knowledge*,
- *properly represent the extracted architectural knowledge*.

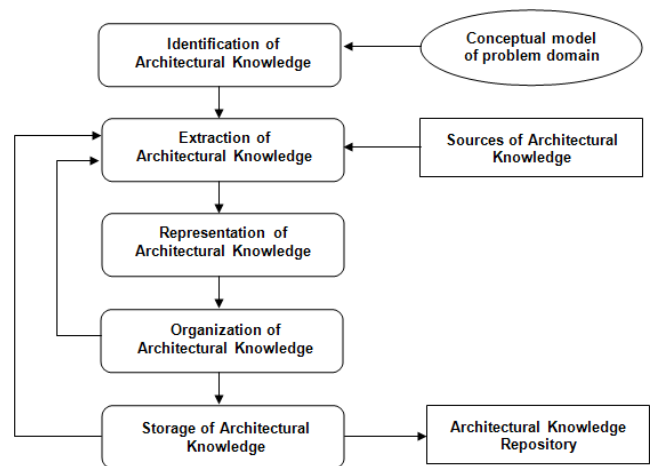


Fig. 1. A conceptual framework for identification and extraction of architectural knowledge

The phase of *identification of the critical architectural knowledge* is aimed at understanding the domain of the problem, identifying the appropriate sources of the architectural knowledge, and deciding about the techniques to be used. Within this phase, the understanding of the problem domain is crucial, as the required architectural knowledge is identified depending on the nature of the problem [5]. So, it can be deduced that the nature or the character of actual problem (e.g. the type and proportion of system's change during the maintenance phase) helps to separate relevant knowledge with regard to presumable purpose of its use.

Other part of this phase involves *identifying the appropriate source of architectural knowledge*, which emphasizes the importance of choosing the adequate way of system's development. *Model-driven architecture* represents one the possibilities, mainly because of its ability to look on the system and its development from various distinct levels of abstraction which helps to avoid the problem of flooding by the high rate of implementation and technological details that are not usually directly related to the planned application of knowledge [10]. Within this approach, the particular models of the system present the primary source of knowledge, so it's inevitable to choose those of them that are sufficient and appropriate sources of architectural knowledge regarding the nature of the problem [8], [9].

The main objective of phase that deals with *extraction of architectural knowledge* is to acquire necessary knowledge from the sources identified in previous phase. The architectural knowledge extracted in this phase needs to be



afterwards *organized and validated* for example using one of the common techniques from artificial intelligence [1], [4]. The process of automated extraction of architectural knowledge depends strongly on its sources and nowadays, it often presents one of the most challenging problems related to knowledge use in software engineering [5]. Other serious problems usually arise while deciding about the character of knowledge representation.

The *choice of an appropriate representation of architectural knowledge* presents an important aspect as it widely influences its further use, processing and evolution.

Knowledge representation is defined as the combination of knowledge structure based on some known representation technique and its interpretation underlain by knowledge languages, usually based on formal logics [1], [4], [10]. While considering the most advantageous combination, it is necessary to take into account the aspect of *integration* of the extracted knowledge into the system's architecture [9].

Current trend promotes the use of *ontologies* as a main representation technique supported by interpretation languages based on description logics, such as *OWL*. This trend is understandable as ontologies present one of the best knowledge representation technique concerning the integration aspect and is currently widely supported by the popular Semantic Web initiatives [1], [10].

#### *B. Integration of architectural knowledge into software system as a part of system's architecture*

The approach of integrating knowledge with the software system tries to overcome the problem of current condition that consists in separation of system's architectural knowledge and system itself and which negatively influences the aspects of accessibility, orientation, application and use of this knowledge within the evolution of the system. Integration of architectural knowledge directly into the system's architecture could widely secure the increase of flexibility, reliability and simplicity of the processes related to maintenance phase [5], [8], [9].

The mentioned approach counts on existence of so-called independent *knowledge layer* that contains the repository of architectural knowledge extracted (and properly represented) from specific system [8]. During the run-time, the system would have an access to the knowledge layer and would use the architectural knowledge (i.e. knowledge about itself) for flexible management of changes and modifications during the maintenance phase. The problem of this approach resides in two main aspects [9]:

- *designing the structure of knowledge layer and defining corresponding access strategies,*
- *analyzing and choosing the most advantageous type of software architecture which would support the integration of knowledge layer.*

While considering the style of software architecture, it is necessary to put the strong accent on the nonfunctional requirements such as *maintainability, integrability, scalability and adaptability*. Whereas there is a requirement of independency of the relation between the base software architecture and the knowledge layer, the final modified software architecture has to be necessarily *loosely-coupled*.

Current trend promotes the application of *service-oriented*

*architecture*, as it presents a flexible, dynamic type of software architecture with extended support of technologies based on the principles of knowledge engineering (e.g. semantic web services, ontologies use, etc.) [11].

## IV. CONCLUSION

In this paper, I presented the principles of an architectural knowledge – a potentially useful way of overcoming the current problems within phases of software life cycle, mainly software maintenance and software evolution. As the concept of architectural knowledge is very new, there are still plenty of obstacles that inhibit it from its further use. I tried to emphasize the most serious problems that exist in this domain by analyzing its partial aspects in a detailed way and to outline some potential solutions that would help the integration of architectural knowledge into the software architecture in order to cope with the problems related to realization of permanent changes within software life cycle.

## ACKNOWLEDGMENT

The paper was prepared within the project "Life cycle and architectures of program systems based on the knowledge", No. 1/0350/08, 2008-01-01 2010-12-31 with the support of VEGA.

## REFERENCES

- [1] H. Happel, S. Seedorf, „Applications of Ontologies in Software Engineering“, Proceedings from 2nd International Workshop on Semantic Web Enabled Software Engineering, 2006.
- [2] H. Yang, M. Ward, „*Successful Evolution of Software Systems*“, Artech House Publishers, 2003, 300 p. ISBN 1580533493.
- [3] K. G. Canfora, A. Cimitile, “Software Maintenance. Handbook of Software Engineering and Knowledge Engineering”, volume 1. World Scientific, 2001, ISBN: 981-02-4973-X.
- [4] J. Durkin, „Expert Systems: Design and Development“, Macmillan: New York, 1994. ISBN 0133486400.
- [5] I. Gorton, A. Babar, „Architectural Knowledge Management: Concepts, Technologies, Challenges“, 29th International Conference on Software Engineering - Companion, 2007, pp. 170-171.
- [6] A. Isazadeh, „Software Engineering: Integration“, Journal of Applied and Computational Mathematics, Vol. 3, No. 1., 2004, pp. 56-66.
- [7] M. G. B. Dias, N. Anquetil, M. K. de Oliveira, “Organizing the Knowledge Used in Software Maintenance”, Journal of Universal Computer Science, vol. 9, no. 7, 2003, 641-658.
- [8] I. Adamuščinová, J. Kunštár, Z. Havlice, „Knowledge in Software Life Cycle“, SAMI Proceedings 2009, pp. 153-157, ISBN 978-1-4244-3802-0
- [9] I. Adamuščinová, „Znalosti, ich reprezentácia a využitie v životnom cykle a architektúrach softvérových systémov“, Písomná práca k dizertačnej skúške. KPI FEI TU v Košiciach, 2008, 112 p.
- [10] D. Gasevic, D. Djuric, V. Devedzic, “*Model Driven Architecture and Ontology Development*”, Springer 2006, ISBN: 3-540-32180-2.
- [11] I. Gorton, „*Essential Software Architecture*“, Springer-Verlag: Berlin Heidelberg, New York, 2006. 283 p. ISBN-10 3-540-28713-2.

# 2D Laser Scanner for the Navigation of the Autonomous Vehicle

František BANÍK

Dept. of Electrotechnical, Mechatronic and Industrial Engineering, FEI TU of Košice, Slovak Republic

frantisek@banik.sk

**Abstract**—The autonomous navigated vehicles are applied in present days more than ever before. Interaction instrument for motion in the space of these vehicles is needed. This instrument is also needed for testing the autonomous theories in the laboratory. The article describes two dimensional (2D) laser scanner based on the one dimensional distance laser sensor. The second dimension was attained with the mechanical rotation construction. The model was real constructed and verified in the office space.

**Keywords**—laser scanner, 2D imaging, autonomous vehicle, navigation

## I. INTRODUCTION

The interaction instrument for the autonomous mode of the vehicle is needed. Interaction instruments can be divided in two groups. In the first group are instruments which work with images from digital cameras. In the second group are instruments which work with physical dimensions of the space. For the second group can be used distance sensors based on infrared, ultrasonic [1] or laser [2]. Technology of the sensor is the main limitation of the measurable distance in the space. Indoor vehicles usually need to measure distance up to 10m. Laser sensors without reflector work up to 6m. Three dimensional (3D) laser scanners are available in present days, but for small projects are still too expensive. Two dimensional (2D) laser scanner, described in this article, works with the laser distance sensor (1D). The second dimension was attained with mechanical moving construction.

## II. MECHANICAL CONSTRUCTION

Laser distance sensor can measure distance between vehicle and obstacle, which is only the first dimension of the space. Rotate mechanical construction was designed to achieve the second dimension. Mechanical construction was constructed from stepping motor, incremental sensor, bearings and from the end limit switch. Transmission between laser sensor and non-rotate parts of the vehicle is problem for non-limited rotation of the scanner, because there is a cable with power supply and signals of the laser sensor. Rotation was limited to  $\pm 180^\circ$ . Working life of the cable for power supply and signals is 30000 cable twists. The stepping motor is the starting point of the rotation part of the mechanical construction. Diaphragm coupling has been used for transmission torque from stepping motor to scanner shaft. Incremental sensor has been mounted over the coupling. Zero point of the incremental sensor has been set to mathematical zero of the scanned space image. Two bearings were used for fixation the scanner shaft. On the top of the shaft bracket for the laser sensor has been mounted. Mechanical position of the laser scanner on the vehicle is

important for the next mathematical calculation of the space map, especially if vehicle is moving in the space.

## III. POWER DRIVE

Stepping motor, Fig. 1, has been used for motion of the laser scanner. Scanning in the discrete points requests usage of the stepping motor. Minimal step of the motor has been chosen as main requirement for selection of the motor. Minimal step is an angle, in which motor can change position and stay in this position with nominal torque. Minimal step  $0.72^\circ$  was proposed for the described laser scanner. Space between scanned points on the obstacle is 6.6cm in distance of 5.3m from the vehicle for this proposed minimal step. Scanning all space around the vehicle ( $360^\circ$ ) includes 500 discrete points. Stepping motor was proposed with nominal static binding torque 37Ncm. This torque suffices for holding of the laser sensor. Theoretical acceleration of the motor is  $66000\text{rads}^{-1}$ . Inertial torque of the system is overcoming by motor in the movement between scanning points. Speed of the movement and stabilization in the new position is important for dimensioning of the task time of the control process V. All space scanning ( $360^\circ$ ) speed depends on the task time and on number of the scanned points. Stepping motor has been connected to the bipolar connection and its current is 0.9A.

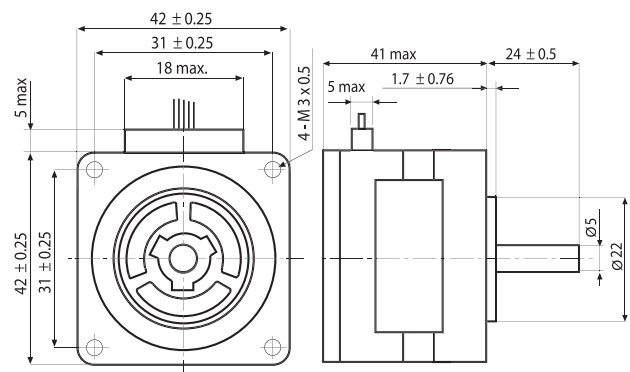


Fig. 1. Stepping motor.

Power converter NDC 04.V has been used for power supply of the motor. Converter receives commands from control unit of the vehicle by three digital inputs. The first one is permission for motor current. Motor current is turned-out in case when exact position is not required. The second digital input is for control of the motor direction. The third signal is pulse coded. Number of the received impulses sets the number of minimal steps, which will be overcome in one movement.

Pulse signals from the incremental sensor have been connected to inputs of the control unit of the vehicle. Absolute position from the incremental sensor is not equal with real position of the laser sensor after power-on of the scanner. Switched-out incremental sensor does not transmit pulse signals and there can be the scanner motion before switch on. For the first time positioning after switch-on the scanner the end limit switch has been mounted. Algorithm of the first time positioning is described in V.

#### IV. LASER SENSOR

The laser sensor SICK DT60, Fig. 3, works on wave length  $650nm$ . Laser beam output is  $\leq 22mW$ . Scanned distance is calculated from the time difference between transmitted and received laser beam. Calculated distance is converted to the electrical signal with range  $4...20mA$ . Output current  $Q_A$  characteristics on scanned distance is shown on figure 2.

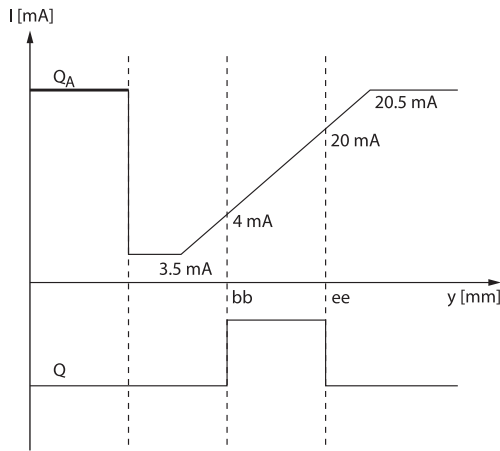


Fig. 2. Output current  $Q_A$  characteristics on scanned distance.

Stabilization time of the output current is  $50ms$ , after change position. Task time of the control process mainly depends on this stabilization time. The laser diode can be switched-off by multifunction input of the sensor. The control unit of the vehicle switches on the laser diode only during scanning space. This function reduces power consumption of the scanner and increases the safety, because the laser diode belongs to category 2 of the norm EN 60 825-1. Covering of the sensor is IP67, which allows usage sensor out of the vehicle cover.



Fig. 3. Laser sensor SICK DT60-211B.

#### V. CONTROL PROCESS

Control process of the laser scanner is one of the tasks of the vehicle control unit. Movement between scanned points and positioning to the start point of the scan is controlled by this process. Period of the task for positioning is  $50ms$ . Period during scanning space is set to  $150ms$ . This time is attained by miss out three tasks with period  $50ms$ . For mechanical and electrical stabilization of the system during the scanning space the extension  $150ms$  is needed.

Exact position of the laser sensor is unknown after power-on of the laser scanner. End limit switch has been mounted in the left shaft stop, as is described in III. Left direction positioning starts after power-on automatically, till the end limit switch is switched-on. Interrupt process ITR1 is linked up to this impulse. The end limit switch signal is approximate information about the laser sensor position. The next step to find the exact position is change direction to right positioning and waiting for finding the zero point of the incremental sensor. Interrupt process ITR2 is linked up to signal from zero point and then incremental sensor counter is set to zero in the control unit of the vehicle. In this point the stepping motor is also stopped and motor current is switched-off. When the power supply of the incremental sensor is on any motion of the laser sensor does not create the lost of the exact position. Laser scanner is ready for scanning space in this state.

The scanning space can be controlled by navigation automatically, or manually in the user mode through the visualization VI. Inputs information before start scanning are start-angle, stop-angle and sampling rate. Start and stop angle can be set in range of  $-180^\circ \dots +180^\circ$ . Sampling rate is the number of stepping motor minimal steps which will be missed between scanning points, range is  $1...3$ . The positioning to the start angle is the first step of scanning, after this the laser diode is turned on. The control unit loads the distance from the laser sensor and calculate angle from counter of the incremental sensor. This data are written down in the matrix 1 in the first row in form  $[\delta, \gamma]$ , where  $\delta$  is the distance and  $\gamma$  is the angle.

$$\begin{bmatrix} [\delta_1, \gamma_1] \\ [\delta_2, \gamma_2] \\ \vdots \\ [\delta_i, \gamma_i] \end{bmatrix} \quad (1)$$

Then control unit sends impulses to the converter for the shift to the next scan point and this task is repeated. Scanning is finished after stop-angle is reached. Result of the scanning is the matrix 1 with the number of rows corresponding to the number of scans, which is index of scanning. The matrix represents orthogonal interpretation of the obstacles in the space. The matrix and index are transferred to the visualization application for processing.

#### VI. VISUALIZATION

The visualization application was developed in development tool CONTROL WEB 5, according to rules described in [3]. Communication between control unit of the vehicle and the visualization is through ethernet with using communication protocol ARION described in [4].

Manual scanning starts by setting input parameters for scanning and then after confirmation the laser scanner realizes process described in V. The orthogonal matrix 1 is transferred

to cartesian coordinate system 3 by equations 2, after receiving the matrix from the control unit.

$$\begin{aligned} x_n &= -\delta_n * \cos(\gamma_n) \\ y_n &= \delta_n * \sin(\gamma_n) \end{aligned} \quad (2)$$

$$\begin{aligned} [x_1, y_1] \\ [x_2, y_2] \\ \vdots \\ [x_i, y_i] \end{aligned} \quad (3)$$

Drawing of the map from the matrix 3 is through the instrument DRAW, which is standard instrument implemented in CONTROL WEB 5. Open polygon with 512 vectors were created in this instrument. Based-points and end-points of this vectors are points from rows of the matrix 3. Vectors of the polygon with the number higher than the index of the scan must be drawn as one point. All 512 vectors can be used only if whole space (360°) has been scanning. Drawn polygon represents vector map of the scanning space in meters. Functions ZOOM, SHIFT and DISCONTINUITY were proposed for the orientation in the map.

### VII. VERIFYING

The laser scanner has been mounted in the model of the autonomous vehicle [5]. The Accuracy of scanning was verified in the office space. Chairs, armchairs, furnitures and paper box were used as obstacles. The first test was performed in the space bounded by this obstacles as is shown in Fig. 4.



Fig. 4. Vehicle position in the office space between obstacles.

Scanning input parameters were: Start angle -180°, Stop angle +180°, Sample rate 1 which is 500 scanning points. The map of the space, drawn in the visualization, is shown in figure 5.

Obstacles can be identified, but there are also vectors of the discontinuity. These discontinuities are searched with the function DISCONTINUITY in the visualization application. The second test was performed in small space because of verifying the map in the photo, Fig. 6.

### VIII. CONCLUSION

The autonomous navigation theories require an interaction element for the testing. The described model of the 2D laser

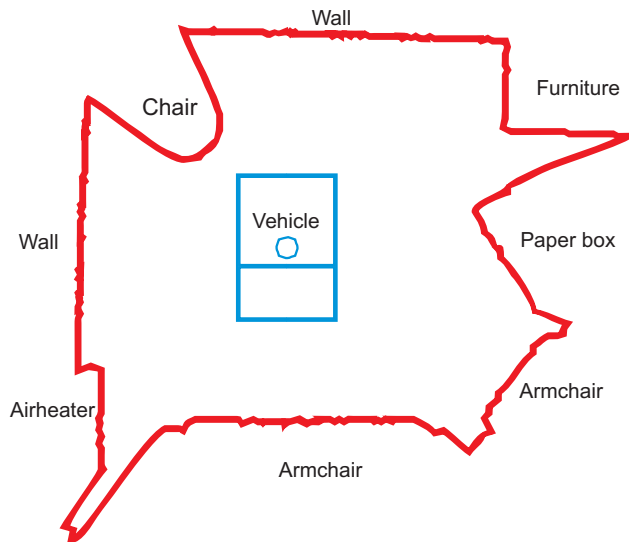


Fig. 5. The map of the office space from figure 4.

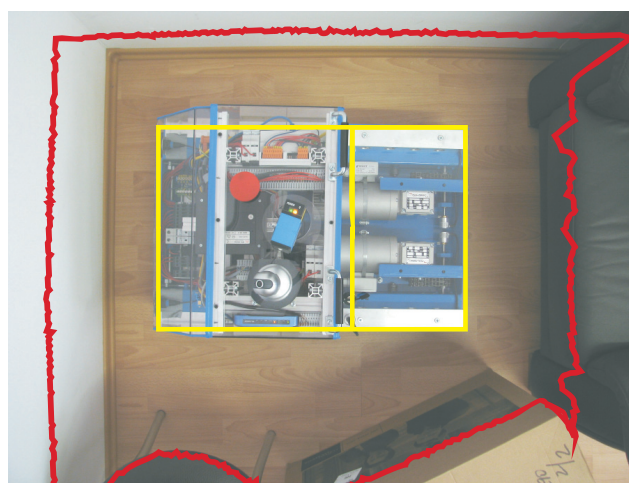


Fig. 6. The map of the small space added to the photography.

scanner is real constructed. Standard automation components were used for construction of this scanner. This attribute allows the usage in the models of vehicle, which dimensions are designed for spaces in offices. The price of this construction is still more cheaper than available commercial 3D scanners. Disadvantage of loss of the one dimension can be accepted in the laboratory testing of the autonomous theories. The scanner was tested in the office space. The results vector maps corresponds with the real dimensions of the space.

### REFERENCES

- [1] S.-Y. Yi and B.-W. Choi, "Autonomous navigation of indoor mobile robots using a global ultrasonic system," *Robotica*, vol. 22, no. 4, pp. 369-374, 2004.
- [2] P. Hoppen, T. Knieriemen, and E. von Puttkamer, "Laser-Radar based Mapping and Navigation for an Autonomous Mobile Robot," *IEEE International Conference on Robotics and Automation 1990*, pp. 948 - 953, May 1990.
- [3] D. Perduková and P. Fedor, *Riadiace a vizualizačné systémy*. TU Košice, 2006, vol. 70, E-Learningový učebný materiál. [Online]. Available: [http://download.ulern.sk/tuke/rvs/ulern\\_viewer.htm](http://download.ulern.sk/tuke/rvs/ulern_viewer.htm)
- [4] L. Frýbort and M. Vosáhl, *ARION Protokol pro ovládání vzdálených V/V*, 1st ed., AMIT spol s.r.o., Chlumova 17, 130 00 Praha 3, Czech Republic, Aug. 2001.
- [5] F. Baník, "Model of the autonomous tracked vehicle," *8th Scientific Conference of Young Researchers of Faculty of Electrical Engineering and Informatics Technical University of Kosice*, vol. 4, no. 1, pp. 118-121, May 2008.

# Software Maintenance

*Peter BRATRU*

Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

peter.bratru@tuke.sk

**Abstract**—Maintenance is the most expensive phase in a software life cycle. This paper briefly defines software maintenance. Then describes three most commonly used traditional process models and five maintenance process models. Program understanding and reverse engineering are as part of the maintenance process in the end.

**Keywords**—software life cycle, software maintenance, process model.

## I. INTRODUCTION

Millions of lines of software are increasing year after year. Software life cycle consists of many phases: analyse, design, implementation, testing. After a product is released, comes the most expensive phase – software maintenance. Maintenance phase keeps the software up to date with environment changes and changing user requirements. It takes life cycle costs up to 80% [2], [3].

## II. SOFTWARE MAINTENANCE

### A. Description

The term software maintenance usually refers to changes that must be made to software after they have been delivered to the customer or user. The definition of software maintenance by [4]:

The Modification of a software project after delivery to correct faults, to improve performance or other attributes, or to adapt the product to a modified environment.

### B. Types of software maintenance

There are four types of software maintenance as it is shown in [2]: corrective, adaptive, perfective and preventive.

Corrective maintenance deals with the repair of faults or defects found. A defect can result from design errors, logic errors and coding errors. Defects are also caused by data processing errors and system performance errors.

Adaptive maintenance consists of adapting software to changes in the environment, such as the hardware or the operating system. The term environment in this context refers to the totality of all conditions and influences which act from outside upon the system.

Perfective maintenance mainly deals with accommodating to new or changed user requirement. Perfective maintenance concerns functional enhancements to the system and activities to increase the system's performance or to enhance its user interface.

Preventive maintenance concerns activities aimed at increasing the system's maintainability, such as updating documentation, adding comments, and improving the modular structure of system.

Among these four types of maintenance, only corrective maintenance is "traditional" maintenance, while the other types can be considered as software "evolution".

## III. THE MAINTENANCE PROCESS

### A. Traditional process models

There are a lot of software process models and their permutations. Here comes an overview of the three most commonly used models.

The code and fix model is model which consists of two phases: writing the code and fixing it. It is very old and not recommended model. The code becomes hard to fix over time and does not give any room for future enhancements.

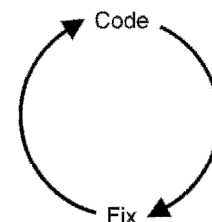


Fig. 1. The code and fix model

The waterfall model gives a high level view of the software life cycle. It is representative model of the well established model and well problem solving mechanism. In contrast to previous model, documentation is an integral part of the process. The problem with this model is that it allows errors in the specification phase, which is more costly to correct at later stage.

The spiral model is iterative model with 4 stages: identification of the objectives, constraints and alternatives, then indentifying the risk among all alternatives. Third step is development of the product and the last planning the next iteration of the spiral. The goal in this model is disclose of risk items. These won not then turn into big issues.

### B. Maintenance process models

Traditional process models are not proper for building maintainability into the software. There are five of the most used models: quick fix model, Boehm's model, Osborne's model, the iterative enhancement model, and the reuse

oriented model.

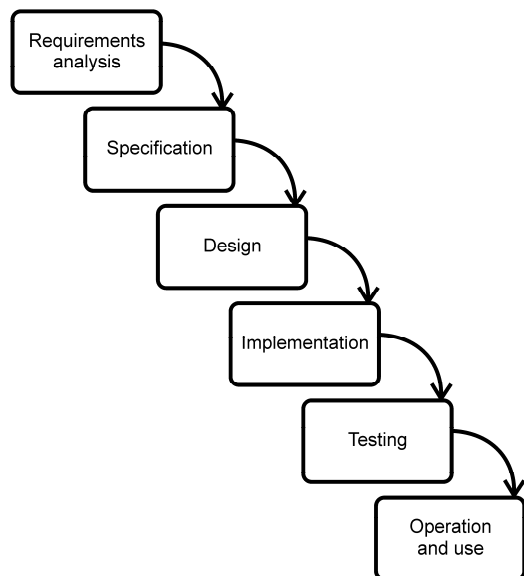


Fig. 2. The waterfall model

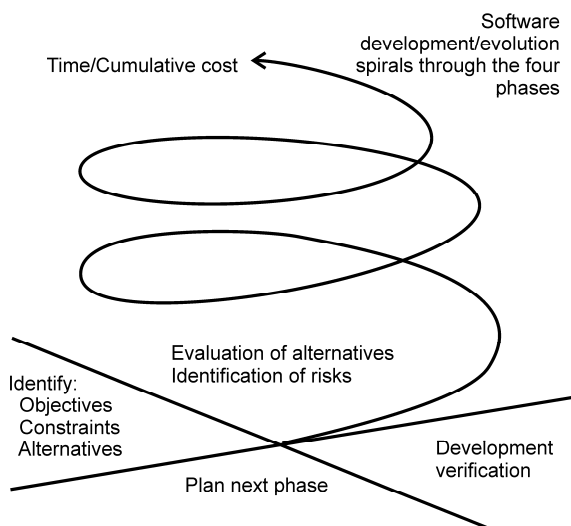


Fig. 3. The spiral model

The quick fix model is an ad-hoc approach. It is the quickest option to fix problem as soon as it is identified. If a system is developed by only one person, that person can make changes without documentation in short time using this model.

Boehm’s model is based on economic models and principles. The use of economic models helps us to better understand the problem and improve productivity in maintenance.

Osborne’s model is concerned with the reality of the maintenance environment. In Osborne’s point of view, technical problems that arise during maintenance are due to poor communication and control between management. Osborne recommends four strategies to address these issues.

1. Maintenance requirements need to be included in the change specification.
2. A quality assurance program is required to establish quality assurance requirements
3. A metrics needs to be developed in order to verify that the maintenance goals have been met.

4. Managers need to be provided with feedback through performance reviews.

The iterative enhancement model can be compared to the traditional spiral model. It is iterative process with 3 stages: system analysis, classification of proposed modifications and implementation. In this model complete documentation of the system is needed for a higher efficiency.

The reuse oriented model assumes that existing program components could be reused. This model has four steps: identifying the parts of the old system which have the potential for reuse, understanding the system parts, modifying the old system parts and integrating the modified parts into the new system.

None of these models is perfect. All models have advantages and disadvantages. Therefore, more than one model is necessary for all maintenance activities [2].

C. Program understanding

Program understanding means having the knowledge of what the software system does, how it relates to its environment, identifying where in the system changes are to be effected and having an in-depth knowledge of how the parts to be corrected or modified work [1].

The maintenance team consists of managers, analysts, designers, and programmers. Every member of team needs to understand the system. There are strategies that could be used to effectively form a mental model for the members of the team. These strategies are top-down model, the bottom-up/chunking model, and the opportunistic model.

D. Reverse engineering

Reverse engineering is the process of analyzing a system to identify the system’s components and their relationships. The aim is to represent system in high level of abstraction. Reverse engineering is required when the process to understand a software system would take a long time due to incorrect, out of date documentation and the insufficient knowledge of the maintainer of the system.

IV. CONCLUSION

This paper is brief summary of software maintenance. The future work can consists of knowledge integration into a system architecture for improve the maintainability process.

ACKNOWLEDGMENT

This work was supported by VEGA Grant No. 1/0350/08 Knowledge-Based Software Life Cycle and Architectures.

REFERENCES

- [1] Grubb, P.; Takang, A. A., “Software Maintenance: Concepts and practice”, World Scientific, 2003
- [2] Stafford, J., “Software Maintenance As Part of the Software Life Cycle” 2003.
- [3] Yip, S.W.L.; Lam, T., “A software maintenance survey,” *Software Engineering Conference, 1994. Proceedings., 1994 First Asia-Pacific*, pp.70-79, 7-9 Dec 1994.
- [4] “IEEE Standard for Software Maintenance”, *IEEE Std 1219-1993*, 2 Jun 1993.

# Performance of Orthogonal Space-Time Block Codes

Peter Drotár

Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

peto.drotar@tuke.sk

**Abstract**—The design of optimal space-time (matrix) constellations for multiple-input multiple-output (MIMO) wireless systems with quasistatic Rayleigh flat-fading channel and coherent maximum likelihood (ML) decoder is an open problem of great interest. Space-time block coding is frequently used method for MIMO communication, with stress on transmit diversity. Space-time block codes (STBC) from orthogonal designs have two advantages, namely, fast maximum-likelihood (ML) decoding and full diversity. In this paper, we compare different orthogonal space-time block codes (OSTBC) for three and four transmit antennas and we will show how is increased overall system performance when additional antennas are added at transmitter.

**Keywords**—MIMO, space-time block codes, orthogonal space-time block codes.

## I. INTRODUCTION

Severe attenuation in a multipath wireless environment makes it extremely difficult for the receiver to determine the transmitted signal unless the receiver is provided with some form of diversity, i.e., some less-attenuated replica of the transmitted signal is provided to the receiver. In some applications, the only practical means of achieving diversity is deployment of antenna arrays at the transmitter and/or the receiver [1].

STBC are the simplest type of spatial temporal codes that exploit the diversity offered in systems with several transmit antennas. A pioneering work in the area of space-time block coding for MIMO wireless channels has been done by Alamouti in [2], in which proposed his space-time block coding scheme for two transmit and multiple receive antennas. Tarokh et al. [1] further generalised the scheme, to include an arbitrary number of transmit antennas by applying the theory of orthogonal designs, thus leading to the concept of OSTBC. The codes developed in [1] are able to provide full transmit diversity specified by the number of the transmit antennas  $N$ , while allowing a very simple maximum-likelihood decoding algorithm, based only on linear processing of the received signals.

Examples of rate 3/4 complex orthogonal designs for  $N = 3$  and  $N = 4$  transmit antennas have appeared in [1], [3], [4], [5]. It was proved in [1] that  $4 \times 4$  complex orthogonal designs of rate 1 do not exist and a simpler proof was given in [3] by using the amicable design theory. This implies that the only square complex orthogonal design of rate 1 is the  $2 \times 2$  complex orthogonal design proposed in [2].

It is important to remember that STBC based on orthogonal design do not achieve a rate of 1 for complex signal constellations. As it was mentioned, it has been shown for 3 and 4

transmit antennas the maximum possible rate is 3/4 with 4 delays. For 5 to 8 transmit antennas, the achievable rate is 1/2 with 8 delays, and for the 9 to 16 case, the rate becomes 5/16 in 16 time instances. In order to achieve the rate of a SISO system, the orthogonal property of STBCs must be broken as described in [6],[7].

The paper is organised as follows. Section II describes MIMO system model which is considered consistently throughout the paper. A short review of space time block coding is outlined in Section III and space time block decoding in Section IV. In the Section V the performance properties of different STBC will be studied by using a set of computer simulations. Finally, conclusions and final remarks to this contribution are drawn in Section VI.

## II. MIMO SYSTEM MODEL

We consider a communication system, where  $N$  signals are transmitted from  $N$  transmitters simultaneously. The signals are the inputs of a MIMO channel with  $M$  outputs. Each transmitted signal goes through the wireless channel to arrive at each of the  $M$  receivers. In a wireless communication system with  $M$  receive antennas, each output of the channel is a linear superposition of the faded versions of the inputs perturbed by noise. Each pair of transmit and receive antennas provides a signal path from the transmitter to the receiver. The coefficient  $\alpha_{n,m}$  is the path gain from transmit antenna  $n$  to receive antenna  $m$ . Figure 1 depicts a baseband discrete-time model for a flat fading MIMO channel. Based on this model, the signal  $r_{t,m}$  which is received at time  $t$  at antenna  $m$ , is given by [8]

$$r_{t,m} = \sum_{n=1}^N \alpha_{n,m} c_{t,n} + \eta_{t,m}, \quad (1)$$

where  $c_{t,m}$  is the noise sample of the receive antenna  $m$  at time  $t$ . Based on 1 replica of the transmitted signal from each transmit antenna is added to the signal of each receive antenna [8].

To form a more compact input-output relationship, we collect the signals that are transmitted from  $N$  transmit antennas during  $T$  time slots in a  $T \times N$  matrix,  $\mathbf{C}$ , as follows:

$$\mathbf{C} = \begin{pmatrix} c_{1,1} & c_{1,2} & \cdots & c_{1,N} \\ c_{2,1} & c_{2,2} & \cdots & c_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ c_{T,1} & c_{T,2} & \cdots & c_{T,N} \end{pmatrix}. \quad (2)$$

Similarly, we construct a  $T \times M$  received matrix  $\mathbf{r}$  that includes all received signals during  $T$  time slots:

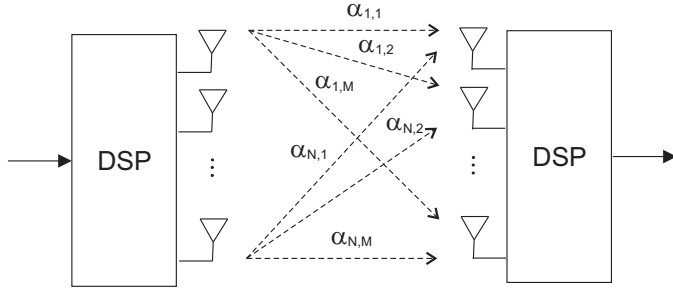


Fig. 1. MIMO block model.

$$\mathbf{r} = \begin{pmatrix} r_{1,1} & r_{1,2} & \cdots & r_{1,M} \\ r_{2,1} & r_{2,2} & \cdots & r_{2,M} \\ \vdots & \vdots & \ddots & \vdots \\ r_{T,1} & r_{T,2} & \cdots & r_{T,M} \end{pmatrix}. \quad (3)$$

Assuming that the fades do not change during the transmission of one block of data and the values of path gains in  $\mathbf{1}$  are constant for every frame we can write  $N \times M$  channel matrix  $\mathbf{H}$  as

$$\mathbf{H} = \begin{pmatrix} \alpha_{1,1} & \alpha_{1,2} & \cdots & \alpha_{1,M} \\ \alpha_{2,1} & \alpha_{2,2} & \cdots & \alpha_{2,M} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{N,1} & \alpha_{N,2} & \cdots & \alpha_{N,M} \end{pmatrix}. \quad (4)$$

Now, the equation  $\mathbf{1}$  can be written in matrix form as:

$$\mathbf{r} = \mathbf{C} \cdot \mathbf{H} + \mathcal{N}, \quad (5)$$

where  $\mathcal{N}$  is  $T \times M$  matrix of additive white Gaussian noise samples [8].

### III. SPACE-TIME BLOCK CODING

The Alamouti scheme can be regarded as a space-time block code with complex signals for two transmit antennas. The transmission matrix is represented by

$$\mathbf{C} = \begin{pmatrix} s_1 & s_2 \\ -s_2^* & s_1^* \end{pmatrix}. \quad (6)$$

At time  $t$  the encoder transmits  $s_1$  from antenna one and  $s_2$  from antenna two simultaneously. At the next time slot  $t + T$ , where  $T$  is the symbol duration, the encoder transmits  $-s_2^*$  from antenna one and  $s_1^*$  from antenna two, where  $s^*$  is the conjugate of  $s$ . The coding delay incurred is equal to  $2T$  [9].

The Alamouti scheme is unique in that it is the only space-time block code with an  $N \times M$  complex transmission matrix to achieve the full rate. If the number of the transmit antennas is larger than two, the code design goal is to construct high-rate complex transmission matrices  $\mathbf{C}_{N,T}$  with low decoding complexity that achieve the full diversity.

Authors in [10] proposed OSTBC for 3 and 4 transmit antennas. Transmission matrix of these codes are as follows:

$$\mathbf{C}_{3,8} = \begin{pmatrix} s_1 & s_2 & s_3 \\ -s_2 & s_1 & -s_4 \\ -s_3 & s_4 & s_1 \\ -s_4 & -s_3 & s_2 \\ s_1^* & s_2^* & s_3^* \\ -s_2^* & s_1^* & -s_4^* \\ -s_3^* & s_4^* & s_1^* \\ -s_4^* & -s_3^* & s_2^* \end{pmatrix} \quad (7)$$

and

$$\mathbf{C}_{4,8} = \begin{pmatrix} s_1 & s_2 & s_3 & s_4 \\ -s_2 & s_1 & -s_4 & s_3 \\ -s_3 & s_4 & s_1 & -s_2 \\ -s_4 & -s_3 & s_2 & s_1 \\ s_1^* & s_2^* & s_3^* & s_4^* \\ -s_2^* & s_1^* & -s_4^* & s_3^* \\ -s_3^* & s_4^* & s_1^* & -s_2^* \\ -s_4^* & -s_3^* & s_2^* & s_1^* \end{pmatrix}. \quad (8)$$

Note, that mentioned transmission matrix are created from  $8 \times 16$  matrix by algorithm described in [10].

In [11] was introduced simpler design of OSTBC for 3 and 4 antennas described by transmission matrix:

$$\mathbf{C}_{3,8} = \begin{pmatrix} s_1 & s_2 & s_3 \\ -s_2^* & s_1 & 0 \\ s_3^* & 0 & -s_1^* \\ 0 & s_3^* & -s_2^* \end{pmatrix} \quad (9)$$

for 3 antennas and

$$\mathbf{C}_{4,8} = \begin{pmatrix} s_1 & s_2 & s_3 & 0 \\ -s_2^* & s_1 & 0 & s_3 \\ s_3^* & 0 & -s_1^* & s_2 \\ 0 & s_3^* & -s_2^* & s_1 \end{pmatrix} \quad (10)$$

for 4 transmit antennas.

Two previously introduced codes characterized by transmission matrix  $\mathbf{C}_{3,8}$  and  $\mathbf{C}_{4,8}$  are obtained from 8 transmission matrix :

$$\mathbf{C}_{8,8} = \begin{pmatrix} s_1 & s_2 & s_3 & 0 & s_4 & 0 & 0 & 0 \\ -s_2^* & s_1 & 0 & s_3 & 0 & s_4 & 0 & 0 \\ s_3^* & 0 & -s_1^* & s_2 & 0 & 0 & s_4 & 0 \\ 0 & s_3^* & -s_2^* & s_1 & 0 & 0 & 0 & s_4 \\ s_4^* & 0 & 0 & 0 & -s_1^* & -s_2^* & -s_3^* & 0 \\ 0 & s_4^* & 0 & 0 & -s_2^* & -s_1 & 0 & -s_3 \\ 0 & 0 & -s_4^* & 0 & -s_3^* & 0 & s_1 & s_2 \\ 0 & 0 & 0 & s_4^* & 0 & -s_3^* & -s_2^* & s_1^* \end{pmatrix} \quad (11)$$

It is possible to obtain codes for  $N = 5, 6, 7, 8$  transmit antennas from transmission matrix  $\mathbf{C}_{8,8}$ , by taking the first  $N$  columns of the introduced matrix.

### IV. DECODING OF STBC

Now let us consider the decoding algorithm. The decoding of the STBC described above can be easily deduced from the encoding matrix. Let us assume that we wish to estimate symbols  $\tilde{s}_k$  and that we have defined by  $r_{t,m}$  the received signal at antenna  $m$  at time instance  $t$ . Let  $\epsilon_t$  denote the permutations of the symbols from the first row to the  $t$ -th row. The column position of  $s_k$  in the  $t$ -th row is represented by  $\epsilon_k$  and the sign of  $s_k$  in the  $t$ -th row is denoted by  $sgn_t(k)$ . Then values to be added at the linear combiner are as follows [12]

$$\tilde{s}_k = \sum_{t=\mu(k)} \sum_{m=1}^M sgn_t(k) \tilde{r}_{t,m} \tilde{h}_{\epsilon_k, m} \quad (12)$$

where  $k = 1, 2, \dots, p$  and  $\mu(i)$  is the set of columns of the transmission matrix, in which  $s_k$  appears. In equation (12)



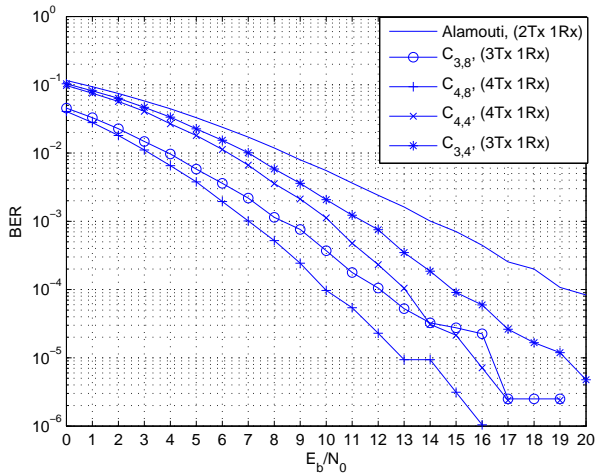


Fig. 2. Bit error probability plotted against  $E_b/N_0$  for orthogonal STBC for 3 and 4 transmit antennas, 1 receive antenna

$$\tilde{r}_{t,m}(k) = \begin{cases} r_{t,m} & \text{if } s_k \text{ belongs to the } t\text{-th row of } \mathbf{C}_{N,T} \\ r_{t,m}^* & \text{if } s_k^* \text{ belongs to the } t\text{-th row of } \mathbf{C}_{N,T} \end{cases} \quad (13)$$

and

$$\tilde{h}_{\epsilon_k,m} = \begin{cases} h_{\epsilon_k,m}^* & \text{if } s_k \text{ belongs to the } t\text{-th row of } \mathbf{C}_{N,T} \\ h_{\epsilon_k,m} & \text{if } s_k^* \text{ belongs to the } t\text{-th row of } \mathbf{C}_{N,T} \end{cases} \quad (14)$$

## V. PERFORMANCE OF STBC

In this section, we show simulation results for the performance of STBC on Rayleigh fading channels. In the simulations, it is assumed that the receiver knows the perfect channel state information. QPSK modulation was used for all simulations. It should be noted, that transmitted power is normalised to the transmitted power of single input single output system.

Figure 2 provides simulation results for two, three, and four transmit antennas. The figure shows bit error probability versus the received  $E_b/N_0$ . We compared codes introduced in [11] and [10]. Alamouti code is depicted for consideration. As can be seen codes from [10] shows 3dB performance gain over codes introduced in [11]. However, these codes have only 1/2 code rate comparing to 3/4 code rate of [11]. Another advantage is lower decoding delay of [11] resulting from shorter block length.

We now examine the performance of a situation with 3, 4, 5, 6, 7, 8 transmit antennas and STBC obtained from transmission matrix  $\mathbf{C}_{8,8}$  as described in Section III on a Rayleigh fading channel. The simulation results show that increasing the number of transmit antennas can provide a significant performance gain. The increase in decoding complexity for STBC with a large number of transmit antennas is very little due to the fact that only linear processing is required for decoding. In order to further improve the code performance, it is possible to concatenate an outer code, such as trellis or turbo code, with an STBC as an inner code [12].

## VI. CONCLUSION

In this contribution, we compared several STBC from literature. It was shown that choosing better code is not simple

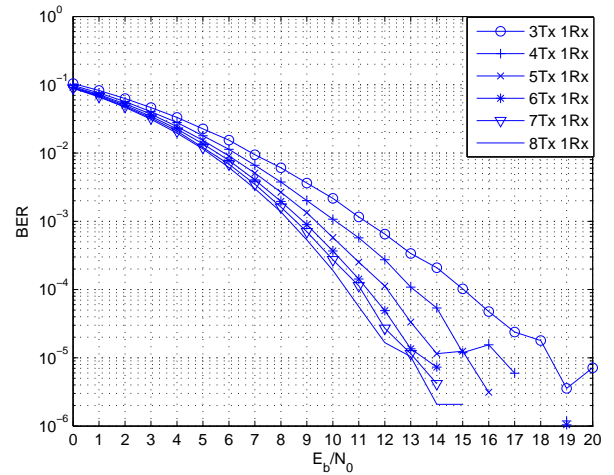


Fig. 3. Bit error probability plotted against  $E_b/N_0$  for orthogonal STBC for 3, 4, 5, 6, 7, 8 transmit antennas, 1 receive antenna

procedure and designer has to take in consideration many code features. Further, we note that an increase in transmit diversity improves the performance. This is a very important inference from a commercial point of view, because handled mobiles always pose a lot of problems in achieving antenna diversity at the receiver.

## ACKNOWLEDGMENT

This work has been funded by VEGA 1/4088/07 "Rekonfigurovateľné platformy pre širokopásmové bezdrôtové telekomunikačné siete", COST 297: "High Altitude Platforms for Communications and other Services" and "Nadácia Tatra banky".

## REFERENCES

- [1] V. Tarokh, H. Jafarkhani and A. R. Calderbank, "Space-Time Block Codes from Orthogonal Designs," *IEEE Trans. on Information Theory*, Vol. 45, July 1999, pp. 1456-1467
- [2] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. on Selected Areas in Communications*, Vol. 16, No. 8, October 1998, pp. 1451-1458.
- [3] G. Ganesan, P. Stoica, "Space-time block codes: a maximum SNR approach", *IEEE Trans. on Inf. Theory*, Vol. 47, May 2001, pp. 1650-1656
- [4] B. M. Hochwald, T. L. Marzetta, and C. B. Papadias, A transmitter diversity scheme for wideband CDMA systems based on spacetime spreading, *IEEE J. on Selected Areas in Communications*, vol. 19, Jan. 2001, pp. 4860
- [5] O. Tirkkonen and A. Hottinen, Square-matrix embeddable spacetime block codes for complex signal constellations, *IEEE Trans. Inform. Theory*, vol. 48, Feb. 2002., pp. 11221126
- [6] A. Boariu and D.M. Ionescu, A class of nonorthogonal rate-one space-time block codes with controlled interference, *IEEE Transactions on Wireless Communications*, Vol. 2, No. 2, March 2003, pp. 270395.
- [7] G. Tsoulos, *MIMO System Technology for Wireless Communications*, CRC Press, Taylor & Francis Group, 2006
- [8] H. Jafarkhani, *Space-Time Coding : Theory and Practice*, Cambridge University Press, 2005, USA
- [9] A. Slaney and Y. Sun, "Space-Time coding for wireless communications: an overview", *IEE Proceedings-Communications*, Vol. 153, August 2006, pp. 509-518
- [10] V. Tarokh, H. Jafarkhani and A. R. Calderbank, "Space-time block coding for wireless communications: performance results", *IEEE J. on Selected Areas in Communications*, Vol. 17, March 1999, pp. 451-460
- [11] W. Su and X. G. Xia, "Two Generalized Complex Orthogonal Space-Time Block Codes of Rates 7/11 and 3/5 for 5 and 6 Transmit Antennas," *IEEE Trans. on Information Theory*, Vol. 49, Jan. 2003, pp. 313-316
- [12] B. Vucetic and J. Yuan, *Space-Time Coding*, John Wiley & Sons, 2003, England

# Automatic Web Service Composition

Zoltán Ďurčík

Dept. Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

zoltan.durcik@tuke.sk

**Abstract**— Web service composition is very discussed in these days. Web services are programs, which are located on networks. This web services may be used by user, or by other web services, thought the use of standard network protocols. Web services communicate by XML messages, most frequently by SOAP message. Every web service provides some functionality. If isn't possible fulfill the request by one web service, there is potential use web service composition. Using the planning method from artificial intelligence is a very good choice for web service composition. The planning techniques which may be use for web service composition are for example: graph oriented planning, heuristic planning, planning by using logical programming.

**Keywords**— web service, WSDL, OWL-S, composition, AI planning, planner, framework

## I. INTRODUCTION

Web services (WS) are distributed programs, which are located on networks (most frequently on internet) and are used by standard protocol (most frequently by HTTP protocol) [1]. Web services communicate by their users and by other web services through Internet by XML messages. Operations description, WS attributes and messages format are accessible by using the Web service interface. The primary key to understanding web services is to understand standards and protocols, with which web services work. Among most fundamental standards and protocols belong WSDL - Web Service Description Language, SOAP - Simple Object Access Protocol and UDDI - Universal Description, Discovery and Integration.

WSDL is XML base description language. WSDL is used for description of web services and has two main aims. First aim is describe web service and second aim is to localize web service. SOAP is XML based protocol. SOAP is used to exchange information through networks, most often by HTTP. SOAP is communication protocol, which is used to allow communication for web services through internet. UDDI is a standard to registration, categorization and searching the web services. The method of working with UDDI remind catalogue, in which are located information about WS providers and the web services which providers provide. OWL (Web Ontology Language) is used to publication and sharing ontologies and appears from RDF (Resource Description Framework). RDF is a system to description an internet resources. This description consists of subject-predicate-object triplet. OWL-S (Web service ontology) appears from OWL and it is ontology to web services descriptions. OWL-S web service description consists of service profile, which describes what web service makes and what functions provides, from

process model, which describe how is user or other web service able to communicate with given web service, and from service grounding, which specifies the ground property of web service, as are communication protocols, messages format, port number etc.

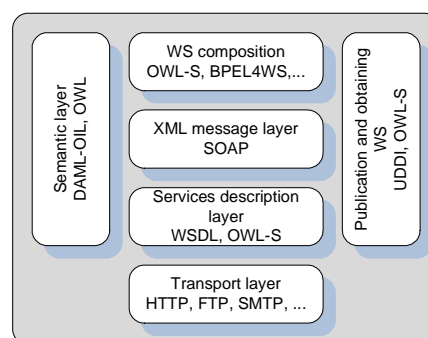


Fig.1. A survey of presented web service technologies

One of the main problems with web services in relation to semantic web is web service composition. Each web service provides some function (e.g. translation words from one language to other). The requirement to composition occur when it isn't possible to achieve desired result by one web service, but it is possible by several web services (e.g. we need to translate one word from Chinese to Slovak, but we don't have directly one web service to this. But there are two web services, which first translate words from Chinese to English and after then second web service translate words from English to Slovak).

Among potentially domain for using web services (and their composition) belongs traveling domain, medical domain, automatic e-mail responding domain, domain for text processing (text-mining), document processing etc.

## II. AUTOMATIC WEB SERVICE COMPOSITION

Generally, more and more organizations, companies and also individuals produce their applications, or provide their services just by web services technologies. Web services have mostly specific inputs and outputs. It means that after our query they provides some information for us on the base of input values. At general holds, that if for our query don't satisfy one web service, it's maybe possible to fulfill this query by combination (composition) more elementary web services. A draft of system for web service composition may be seen on Fig.2.

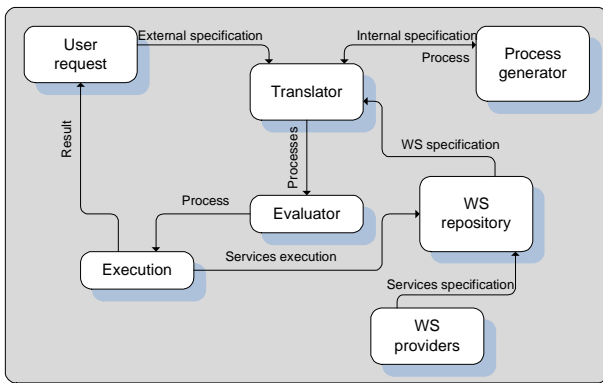


Fig.2. A Framework for web service composition

- *Service repository* - serves to storage atomic web services from WS providers.
- *Translator* - is a component, which is used to processing information obtained from user and from service repository.
- *Process generator* - is a main component of the system. It is an application of algorithms, which are following the user query and the actual state to choice atomic web services, which are able to fulfill this query. Therefore on the end of this process is for us available set of atomic web services together with working and data flow among these services.
- *Evaluator* - evaluation occurs in case, when several plans are generated.
- *Execution engine* - execution selected web services in order by planner. The result is provided to user.

### III. PLANNING BY ARTIFICIAL INTELLIGENCE METHODS

As one of most suitable choice for web service composition is using an artificial intelligence planning methods. Planning problem may be represented as world model, and this model is possible write as pentad:  $\langle S, S_0, G, A, \Gamma \rangle$ .

$S$  represents a set of all possible states in given model,  $S_0$  is a subset of  $S$  and marks a initial state of the world,  $G$  marks goal state,  $A$  is a set of available actions, each of which change world state as transition from one state to another state, and  $\Gamma$  is a subset of  $S \times A \times S$  and define precondition and effects for each action.

A relation between planning and automatic web service composition is following: sets  $S_0$  and  $G$  represent initial and goal state, and these states may be represented for example by ontologies e.g. OWL.  $A$  is a set of actions, and these actions may be represented by available atomic WS.  $\Gamma$  represent a functions for state change for each service. As most suitable language for web service description in relation to artificial intelligence planning has turned out to be OWL-S language. By OWL-S language is possible directly describe in addition to input and output also preconditions and effects.

Among most frequently used artificial intelligence planning techniques belong state-space planning, graph-oriented planning, planning by using hierarchical networks and planning by using logical programming.

#### A. State-space planning

State-space consists from following parts [3]:

- $S$  - enclosed set of possible states,
- $A$  - enclosed set of available actions,
- $f$  - is a function, which describe transitions between individual states,
- $c(a,s) > 0$  - determines "price" of application action  $a$  in state  $s$ .

State-space, which contains descriptions of initial state  $S_0$  and goal state  $S_G$ , is often marked as state model. The aim of state model solution is determine actions sequence  $\{a_0, a_1, \dots, a_n\}$ , which generate a sequence of states  $s_0, s_1=f(s_0, a_0), \dots, s_{n+1}=f(s_n, a_n)$ , where action  $a_i$  is possible use in state  $s_i$ ,  $a_i \in A(s_i)$ , where  $A(s_i)$  is enclosed set of available actions in state  $s_i$  and state  $s_{n+1} \in S_G$  is a goal state of state model.

In general may be used for state-space planning random searching algorithm. But state space reflects a situation from real world may be very extensive. Therefore it is necessary be care at selection of searching algorithm in consideration of his performance and especially then, how it is able to manage with extensity research space.

Among the first attempts to reduce state searching space belonged STRIPS (Stanford Research Institute Problem Solver) algorithm [3]. STRIPS uses backward chaining. It works with following elements:

- Initial state
- Goal state
- A set of actions, where each of these actions includes:
  - o preconditions - they inform about it, what must be fulfilled for execution given action,
  - o effects - they inform about it, what will be change in model after execution given action.

#### B. Graph oriented planning

Graph oriented planning for its function uses graph structures. These graph structures are referred to as planning graphs [5][9]. Planning graph is different of state-space graph, in which states are represented as nodes and edge represented transition between individual states. Planning graph consists from two types of nodes, concretely the nodes of actions and conditional nodes (also marked as prepositional nodes). These nodes are located in alternating layers. The layer of conditional nodes is followed by action nodes layer etc.

#### C. Planning by using hierarchical networks

Relations between individual actions of planning system are expressed by networks. Planning problem is specified in hierarchical task network by following set of permission tasks:

- *primitive task* - respond to actions in STRIPS,
- *composite tasks* - may be composite from several simpler tasks (from primitive tasks or also from another composite tasks),
- *goal tasks* - respond to goal in STRIPS.

Primitive tasks represent actions, which may be directly executed. Composite tasks represent a sequence of actions and a goal tasks represent conditions. Composite tasks for their self execution need know a sequence of primitive tasks, from which they are composite. Goal tasks represent conditions, which must obtain truth value in goal state.

#### D. Planning by using logical programming

By logical programming we consider a program, which may be represented as a set of Horn clauses in a form of implication  $A \leftarrow (B1 \wedge B2 \wedge \dots \wedge Bn)$ .

In a case of relation between logical programming and planning it turned out as suitable methods the methods grounded on deductive consideration, e.g. in a case of PROLOG [10]. Among another application of logical programming belongs e.g. Reiner's implementation of GOLOG and situation calculus [11]. Author is concentrate on knowledge oriented GOLOG programs, which may contain sensing actions. They are designed for on-line execution under the assumption of dynamically encapsulated world. During execution are automatically adapted axioms by following sensing actions.

### IV. A SYSTEMS FOR WEB SERVICE COMPOSITION

#### A. (Semi)automatic web service composition using PROLOG

This system utilizes elements of logical programming, concretely PROLOG language [10]. Prototype of web service composition system has two main components, concrete composition engine (CE) and inference engine (IE).

Inference engine stores information about well-known services in a base of knowledge and in a case of need it is able to find suitable services. Composition engine guard in addition to composition also interaction between users and inference and composition engine.

Inference engine is used to OWL reasoning and it is built on PROLOG. Information describing in OWL are converted to the RDF triplet and saved into base of knowledge (KB). IE has inbuilt an inference rules for OWL. These rules are applied to the fact from KB in order to discovery all dependencies. Such is e.g. discovery inheritance between two classes, which isn't directly encoded in section of relationship between classes.

#### B. SHOP2 planner web service composition

SHOP2 [6] is domain independent HTN planning system, which won one from the four main prizes among 14 planners on International Planning Competition in the year 2002 (IPC-2002). HTN is an artificial intelligence planning method, which is focused on creation plan by task decomposition. Task decomposition is performed so far forth all tasks are not decomposition into primitive tasks.

System architecture for web service composition by using planner SHOP2 is composite from following parts.

- Planner SHOP2 - following initial, goal and available actions create plan, which is described in a SHOP2 domain
- Tool for WS management - has in tasks save information about availability WS and execution plan.
- Tool for representation resultant plan in OWL ontology - provides plan representation in OWL standard.
- Tool for translation OWL-S service description into SHOP2 planning domain - translates OWL-S service description into SHOP2 domain (e.g. atomic processes from OWL-S into HTN planning procedures).

#### C. OWL-SXPlan

OWLS-XPlan [7] is a tool developed to realization semantic web service composition by artificial intelligence planning. It realizes converting web services described by OWL-S 1.1 standard into equivalent problem realized in PDDL<sup>1</sup> 2.1 language. After this problem converting into planning domain is realized planning by using AI planner XPlan. Output from given planner is a web service (action) sequence also referred to as plan. XPlan is base on extension fast forward planner by HTN planning technique.

### V. SUMMARY OF AI PLANNING METHODS AND SYSTEMS

TABLE I  
Comparison selected systems for web service composition

System	Planning methods	WS standard support	Internal problem representation	System type	Dynamically interference during planning
WS composition Prolog	Logical programming	OWL-S WSDL	Logical programming program	Semiautomatic composition with user collaboration	No
Shop2 system	HTN planning	OWL-S WSDL	PDDL 2.1 version 2	Automatic composition	No
OWL-SXplan	Extended Fast forward planning FF and HTN planning	OWL-S WSDL OWL	PDDL 2.1 version 2	Automatic composition	Yes, OWL-Sxplan version 2

A majority of presented AI planning methods have roots in STRIPS. STRIPS may be defined as set  $\langle A, O, I, G \rangle$ , where  $A$  is a set of conditions,  $O$  is a set of operations (actions),  $I$  is an initial state and  $G$  is a goal state in a planning model. Planning objective is find the way, which obtain system (model) from initial state  $I$  into goal state  $G$  by using operators from  $O$  and by compliance conditions from  $A$ . Here is clearly visible analogy with web service composition. In a case of composition it works also with planning. Initial state is a state, in which we are at the start of composition. Goal state is a state, in which we will be at the end of composition. Conditions result from domain, in which we make composition, and from conditions, which are given for individual web services. Web services represent actions. Therefore after appropriate modification and transformation is possible use for web service composition just artificial intelligence planning methods. Given transformations are e.g. transformation initial and goal state, which may be described by OWL ontologies, and web service transformation described by OWL-S language, into PDDL language. This solution was presented e.g. in OWL-SXPlan system.

### VI. SYSTEM ARCHITECTURE FOR WS COMPOSITION

On Fig.3. is presented a system architecture, which should have our planning system. Currently is it only architecture draft and individual block of this system can be modifying eventually into final version. This final version will be implemented.

<sup>1</sup> PDDL (Planning Domain Definition Language) is an attempt to standardization descriptive languages for planning domain and problems. It was developed for ICP (International Planning Competition) 1998/2000.

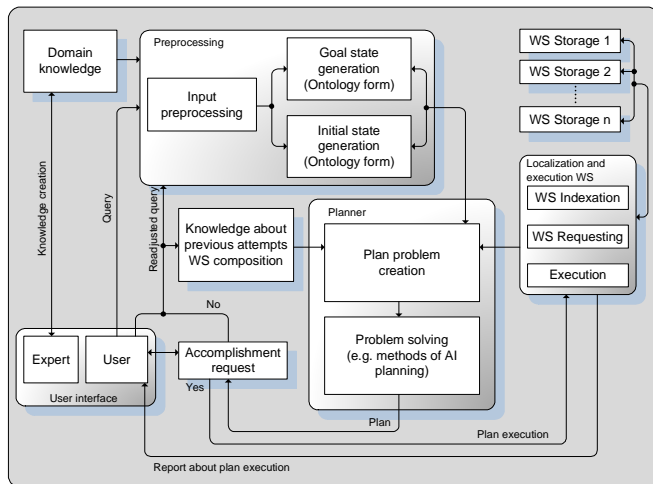


Fig.3. System design for web service composition

Presented system draft consists of following four main parts:

1. *User interface* - provides interaction between users and system. This section may be divided into two parts:
  - a. *user part* - it serves for ordinary user, with which is system used,
  - b. *expert part* - serves for expert user, which is able to modify system function, e.g. modification domain knowledge.
2. *Preprocessing* - provides initial and goal state creation, which is obtained from user query. We assume, that these two states will be represented by ontologies. This part cooperates by domain knowledge.
3. *Planner* - with following initial state, goal state, available web services and information about previous composition attempts is created planning problem and planning domain. This model and problem is next solved.
4. *Localization and execution web services* - has in task interaction with WS, WS indexation and WS execution.

## VII. CONCLUSION - CURRENT PROBLEMS WITH AUTOMATIC WEB SERVICE COMPOSITION

A complicity control is one from problem current systems for web service composition. Correct described initial and goal state are important for result and full process of planning. It is important note that the creation of fully automatic composition system is very complicated, and whereby more we shall be try make system more automatically, thereby more and more problem may stand forth. E.g. system will be increasingly specialized to only some field of solving problems, or system often will not provide relevant result for us etc. Therefore could be suitable compose in system also the possibility system interaction with user during planning. For example in OWL-SXplan system is possible to user dynamically interfere into plan during planning. This hits from user but aren't saved into system and in a case when is produced the same plan again, the system will be work quite the same. Here could be suitable try saved this obtained knowledge from user interference during planning.

Ideal system for automatic web service composition should contain following parts:

- it should provide possibility of addition domain ontologies for planning problem,
- the possibility of searching and indexing WS, or simple possibility to registration WS by users,
- should be for user simple describe initial and goal state following ontologies. In a ideal case could obtain initial and goal state what most automatically (e.g. at some domain could be able obtain some information from user profile),
- should have include efficient tool for web service composition. As more suitable are showing AI planning methods,
- should make possibility dynamically interfere in planning process with user. Suitable representation of planning process is representation by graph structures. In a case of need should user have view how actual planning proceeding and he might interfere into its following self decision,
- system should show intelligently. Intelligence could be inhere e.g. in saving knowledge from user interference during planning.

## ACKNOWLEDGMENT

The work presented in this paper was supported by the following projects: the Slovak Research and Development Agency under the contract Nr. APVV-0391-06; the Slovak Grant Agency of Ministry of Education and Academy of Science of the Slovak Republic under grant Nr. 1/4074/07.

## REFERENCES

- [1] WS: Web Services Glossary, <http://www.w3.org/TR/wsa-reqs/>, 2002
- [2] RAO, J. – SU, X.: A Survey of Automated Web Service Composition Methods. In Proceedings of the First International Workshop on Semantic Web Services and Web Process Composition, SWSWPC 2004, San Diego, California, USA, 2004.
- [3] NILSSON, Nils J. – FIKES, Richard E.: STRIPS: A New Approach to the Application of Theorem Proving to Problem Solving, Artificial Intelligence, 2(3):189-208, 1971.
- [4] WU, Dan – SIRIN, Evren – HENDLER, James – NAU, Dana – PARSIA, Bijan: Automatic Web services composition using SHOP2, In Workshop on Planning for Web Services, 2003
- [5] PEER, Joachim: Web Service Composition as AI Planning - A Survey, Dissertation, University of St. Gallen, Switzerland, 2005.
- [6] SIRIN, Evren – BIJAN, Parsia – WU, Dan - HENDLER, James: HTN planning for web service composition using SHOP2, Journal of Web Semantics 1 (4), pp.377-396, 2004
- [7] KLUSCH, M. - GERBER, A. - SCHMIDT, M.: Semantic Web Service Composition Planning with OWLS-Xplan.1st Intl. AAAI Fall Symposium on Agents and the Semantic Web, Arlington VA, USA, 2005
- [8] OH, S. - LEE, D. - KUMARA, S. R.: A comparative illustration of AI planning based web services composition. ACM SIGecom Exch. 5, 5 Jan., 2006
- [9] BLUM, Avrim L. – FURST, Merrick L.: Fast planning through planning graph analysis, In Artificial Intelligence journal volume 90, 1636 -1642, 1997
- [10] SIRIN, Evren – HENDLER, James – BIJAN, Parsia: Semi-automatic composition of web services using semantic descriptions in Web services: Modeling, Architecture and Infrastructure workshop in ICEIS 2003, Angers, France, April 2003
- [11] REITER, Ray: On knowledge-based programming with sensing in the situation calculus. ACM Trans. Comput. Logic 2, 4 (Oct. 2001), 433-457, 2001

# Map creation based on camera image

<sup>1</sup>Juraj EPERJEŠI

<sup>1</sup>Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

<sup>1</sup>[juraj.eperjesi@tuke.sk](mailto:juraj.eperjesi@tuke.sk)

**Abstract**—Map creation is basic task in mobile robotics, because map allows robot to move freely in the environment and plan its movement. Creating a map from image information is interesting, because the image then can be used for further processing. This article contains a design of system for map building.

**Keywords**—image processing, map building, mobile robotics.

## I. INTRODUCTION

Autonomous robot must be able to move in the environment freely. To do this, robot must know its precise position in each moment. Such information can be send to the robot, but it decreases its autonomy. If the robot can gain this information through its own sensors, it's the best possible situation. To know its position means, that robot knows the position of all obstacles around him. Obstacles are places where robot cannot go and obstacles represented certain way create the map of the environment. Without it, the robot would know only its closest surroundings and therefore could only decide its actions only in range of its sensors. It couldn't also plan the way to places behind the corner, or get through the maze.

Such map must be created somehow. Robot could get the map of the environment from the user directly, but after some possible change in the environment, this map would be useless. This means, for example, that if in the environment is a certain obstacle which isn't in the map, robot can bump into it without knowing it. If user wants the robot to have the most recent map, he must get it to the robot once again. Another disadvantage of this approach is different way of sensing the environment for the robot and human, which could cause, that the robot can't use map given by the user.

This means that map creation is essential for most mobile robotic tasks. It doesn't have to be the global map of the whole environment, but it must be at least local map of the closest vicinity. Creation consists from two steps. First is image processing, in this case edge detection, followed by actual map creation from found edges based on chosen representation. Representation influences every aspect of work with map, its creation, updating, dynamic obstacles etc. One of the most important parts of system is self localization of robot, which means, that robot must be able to determine its position in the map from the sensor readings. Largest problem occurring in mobile robotics is odometric error, which means, that robot may be in the different place

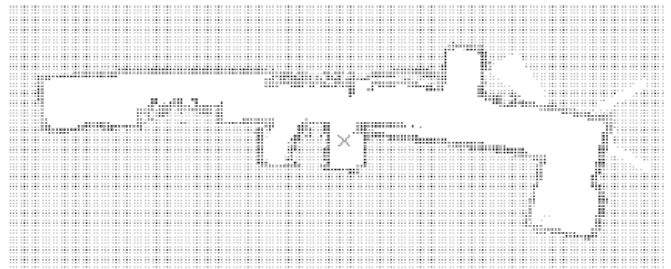


Fig. 1. Map represented as grid with odometric error. [11]

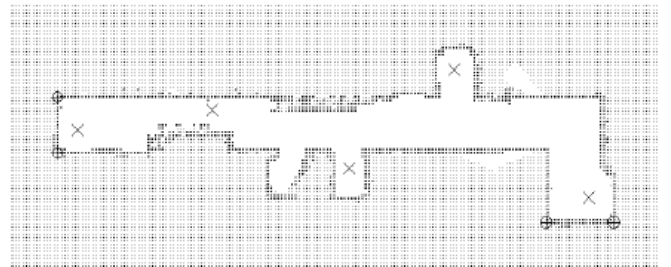


Fig. 2. Same map as in Fig. 1 without odometric error. [11]

according to where he thinks he is. This error is not systematic and can lead to creation of unusable map.

## II. IMAGE PROCESSING

### A. Image Problems

Image gained from the camera must be preprocessed because of three basic problems occurring on camera images. Radial distortion is effect seen near the edges of the image. Problem arouses when system is trying to decide the distance and angle to the certain point. Radial distortion can be repaired, for example OpenCV library for computer vision provides such function. Another problem is vignetting which can be seen in the corners of the image as slight darkening of the image. It is caused by the fact, that the lenses are round and the CCD sensor is rectangular. In [10] this problem was solved by simple brightness normalization based on the distance from the center of the image. This correction is applied for each color independently. Perspective problems emerge because the camera is tilted towards the floor in order to see as close before the robot as possible. They can be omitted in this system, because it observes only edges between obstacles and floor and creates 2D map, but they must be solved if the task would be to create 3D map.

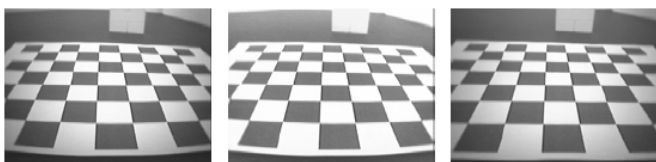


Fig. 3. Image problems. a) original image b) image without vignetting c) image without radial distortion. [10]

### B. Edge Detection

This is the most important part of proposed system. Its output is basis for all consequent steps. In process of choosing right edge detection method, important attribute is speed of detection. The effort is towards real-time detection, or at least as fast detection as possible, so that it can be performed with certain frequency.

Well known methods like Canny [1][12][14] or Sobel [9][13] edge detection were tested. Another edge detection method based on WIR (Window Intensity Regularity) was also tested [6]. Best results were gained with the last method, but also the time was the worst. Basic step is creation of circular window around actual point of interest.



Fig. 4. 1) edge point on the edge between textured and non-textured area, 2) non-edge point inside the textured area, 3) edge point. All points have their windows drawn. [6]

Number of brightness levels is then decreased from standard 256 down to 32 or even 16. Window is then divided into halves horizontally and vertically and for both possibilities histograms are made. After that, Euclidean distances between corresponding histograms are enumerated, and this value serves as input to the Cauchy function. Its output along with some other elements enters this equation

$$E = (1,0 - R(c^2)) \sqrt{\left(\frac{\partial I_s(x,y)}{\partial x}\right)^2 + \left(\frac{\partial I_s(x,y)}{\partial y}\right)^2} + b \sqrt{c_x^2(x,y) + c_y^2(x,y)}$$

where,

$$I_s(x,y) = I(x,y) \cdot G_s(x,y)$$

$$G_s(x,y) = \frac{1}{\sqrt{2\pi}s} e^{-\frac{x^2+y^2}{2s^2}}$$

$R(c^2)$  is Cauchy function

$E$  is then transformed into brightness value and gradient image is created. Size of the window is at least 20 pixels which makes this method very computationally demanding. But its results, especially in textured areas are the best.



Fig. 5. Edge detection by Canny edge detector.



Fig. 6. Edge detection by Sobel edge detector.



Fig. 6. WIR

Result from WIR must be then processed by another edge detection method. But it can be seen, that Canny detects edges on the carpet, which is textured area, Sobel don't, but also omits several important edged, where WIR copes with textures very well while preserving important edges.

### III. MAP BUILDING

In most cases, map is represented by grid, where each cell describes small portion of space, whether there is obstacle or not. This value is mainly the probability of occupancy. In some cases, probability of cell being free is also defined, and each cell then holds two different values. [XXX] Main advantage of this representation is simplicity of use, but is

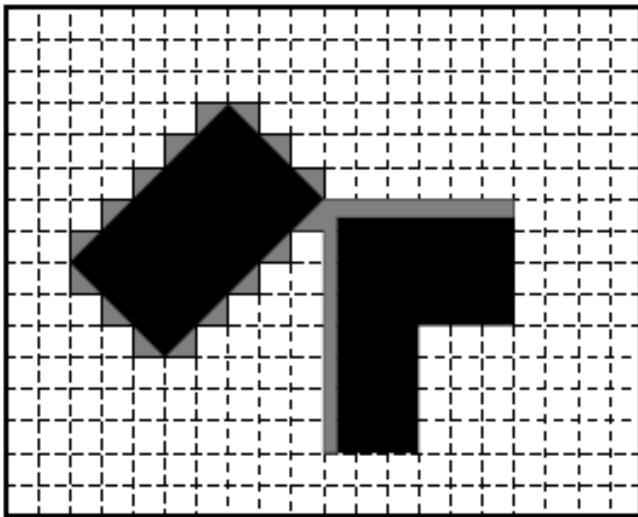


Fig. 7. Difference between real size of objects and their representation in the map. [15]

has also some disadvantages in precision and memory usage.

Occupancy grid can be transformed into topological representation, which is much less demanding in terms of memory usage, but its main disadvantage is, that robot knows its exact location only in specified nodes in the graph. These

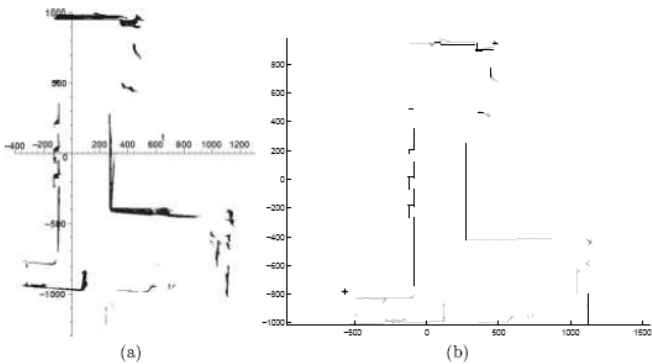


Fig. 8. a) map represented by 144400 points from laser scanner b) map represented by 50 lines in geometric representation. [4]

arguments led to the use of geometric representation in proposed system.

Transformation of points in the image into distance from robot is based on known height of the camera above the ground and angle of camera tilt. One important assumption must be fulfilled, that the robot is on flat floor. Most important points are ones on the corners of obstacles, because these are sufficient for representation of ends of lines. Image quantization brings certain amount of error into calculation of distance and this error grows with distance between point and robot. (See Fig. 9)

This problem was solved in [2] in interesting way, where the map was build not based on exact location of end points of edges, but on angles between edges. Assumption used in this case was, that nearly all angles in interior are 90°.

#### IV. LOCALIZATION

Main problem during map building is that new

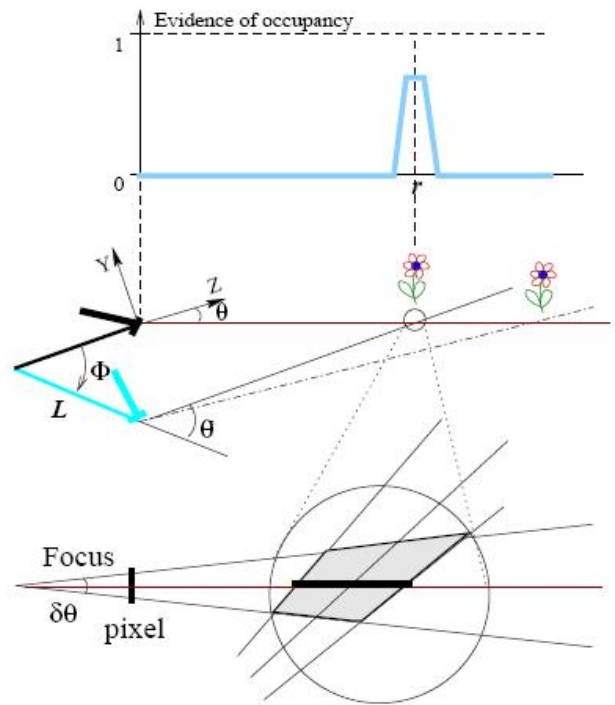


Fig. 9. top – error caused by radial distortion, when one camera observes the same object from two angles and in one case this object is near the edge of image where the image is distorted  
bottom – error caused by size of pixel, exact location can be everywhere in the grey area, in most cases is supposed to be in the middle [7]

information in map must be precise. To achieve this, robot must know its precise location in moment, when map is being updated. But how can robot know its position, when it hasn't the map of environment. This problem is in robotics known as SLAM (Simultaneous Localization and Mapping). In [11] is its solution based on several short-term local maps. With certain frequency, new short-term map was created and the oldest map was destroyed. But before that, this map was used to determine the odometric error. All local maps were corrected and the information from destroyed map was added

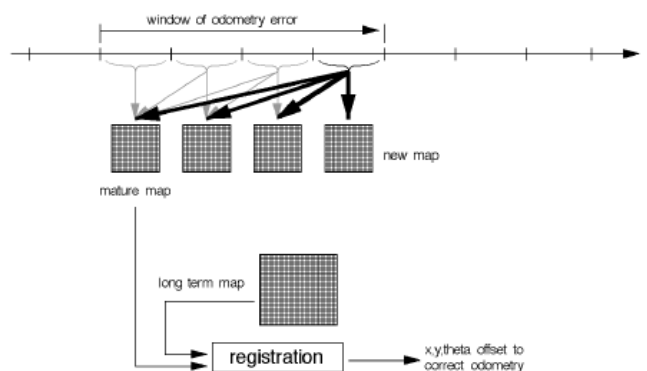


Fig. 10. scheme of using short-term local maps in SLAM problem. [11]

to the global map. Because the frequency of self-localization was rather high in comparison with ordinary approach, supposed corrections were small ( $\pm 15\text{cm}$  and  $\pm 2^\circ$ ) [8].

#### V. MAP CONNECTING

Proposed system for map building is developed as part of larger system MASS [7]. One of main characteristics of MASS is its possibility to manage several agents. That's the



reason why in proposed system is designed part responsible for map connecting. Utilization of such component is in situation where several agents search the same space, or room. In the certain moment, they should realized, that they are in the same room. This should be when their maps begin to overlap. They are then connected and all agents then share this new map.

## VI. ROBOT CONSTRUCTION

Robot is based on LEGO Mindstorm NXT [5], which serves as the chassis for the camera. Robot has 2 driven wheels and communicates with the computer through Bluetooth connection. Main block allows connection of one another motor and 4 different sensors, which gives the robot possibilities for later improvement. Camera used for vision is capable of movement in horizontal and vertical direction with range: pan  $\pm 165^\circ$  and tilt  $-15^\circ - +90^\circ$ . Motors used on camera have higher precision than LEGO motors and therefore it is better to use the camera to look around while robot is standing still. This way, possible odometric error is much lower than it would be if the whole robot was moving. Camera can also connect to the internet through wifi, so

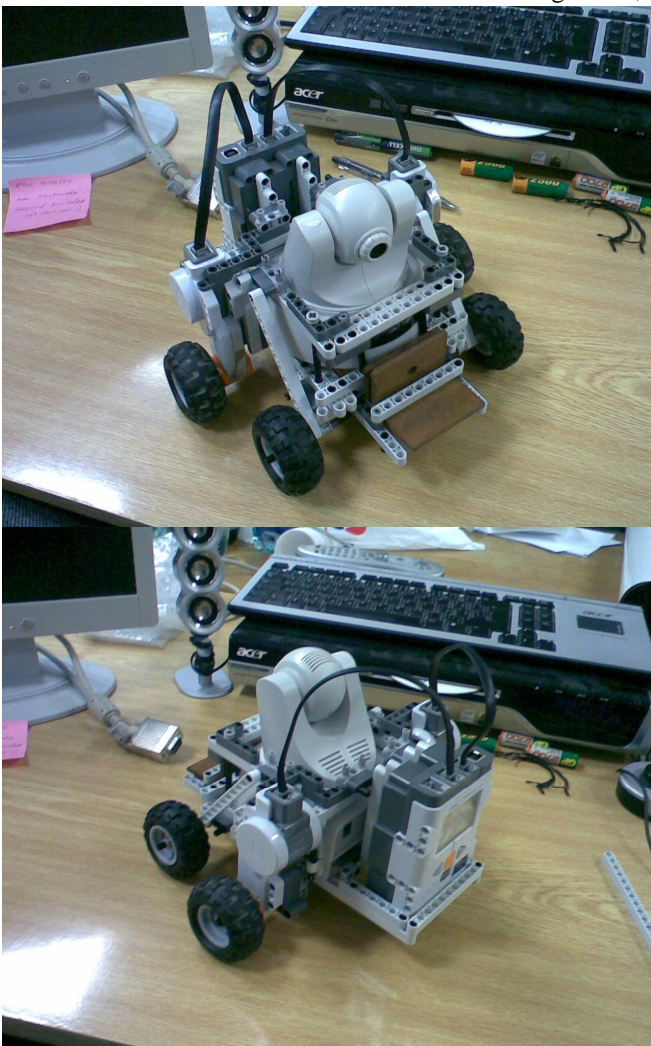


Fig. 11. Construction of robot.

when there is possibility to provide it with the power supply,

robot can be wireless. During experiments, camera is powered from electric outlet and also connected to the internet by cable.

## VII. CONCLUSION

This article described proposed system for map building and pointed out possible ways how to cope with tasks to come. In edge detection area, basic experiments with several methods were described and the construction of robot was presented.

## REFERENCES

- [1] Canny, J.: A Computational Approach to Edge Detection, *IEEE Trans. Pattern Analysis and Machine Intelligence*, 8:679-714, 1986
- [2] Gini G., Marchi A.: Indoor Robot Navigation with Single Camera Vision, In PRIS, ICEIS Press, 2002, pp. 67-76
- [3] Gorodnichy D. O. and Armstrong W. W.: Single Camera Stereo for Mobile Robot World Exploration, In proc. of Vision Interface, 1999, pp. 528-535
- [4] Lakaemper R., Latecki L. J., Wolter D.: Geometric Robot Mapping, <http://www.cis.temple.edu/~latecki/Papers/dgciRobot05.pdf>
- [5] [mindstorm.lego.com](http://mindstorm.lego.com)
- [6] Qixiang Y., Wen G., Weiqiang W.: A New Texture-Insensitive Edge Detection Method, ICICS-PCM 2003, 15-18 December 2003, Singapore
- [7] Reiff, Tomáš: Inkrementálny systém pre rozpoznávanie obrazových vzorov. Diplomová práca. Košice: Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky, 2008. 66 s.
- [8] Schultz A. C. and Adams W.: Continuous Localization Using Evidence Grids, Navy Center for Applied Research in Artificial Intelligence, 1998, pp. 2833-2839, IEEE Press
- [9] Sobel, I., Feldman, G.: A 3x3 Isotropic Gradient Operator for Image Processing, presented at a talk at the Stanford Artificial Project in 1968, unpublished but often cited, orig. in *Pattern Classification and Scene Analysis*, Duda, R. and Hart, P., John Wiley and Sons, 73, pp 271-2
- [10] Taylor T., Geva S., Boles W. W.: Early Results in Vision-Based Map Building, Eds. Proceedings 3<sup>rd</sup> International Symposium on Autonomous Minirobots for Research and Edutainment, 2005, pp. 207-216, Fukui, Japan
- [11] Yamauchi B., Schultz A., Adams W.: Mobile Robot Exploration and Map/Building with Continuous Localization, from Proceedings of the 1998 IEEE International Conference on Robotics and Automation, May 1998, Leuven, Belgium, pp. 3715-3720
- [12] [http://en.wikipedia.org/wiki/Canny\\_edge\\_detector](http://en.wikipedia.org/wiki/Canny_edge_detector)
- [13] [http://en.wikipedia.org/wiki/Sobel\\_operator](http://en.wikipedia.org/wiki/Sobel_operator)
- [14] [http://www.pages.drexel.edu/~weg22/can\\_tut.html](http://www.pages.drexel.edu/~weg22/can_tut.html)
- [15] [www.cosy.informatik.uni-bremen.de/staff/barkowsky/semSpaCog/SpatialRepresentationsForMRN080503.pdf](http://www.cosy.informatik.uni-bremen.de/staff/barkowsky/semSpaCog/SpatialRepresentationsForMRN080503.pdf)

# Classification of isolated words with Point-Border Artmap

Zlatko FEDOR, Tomas REIFF

Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

zlatko.fedor@tuke.sk, tomas.reiff@tuke.sk

**Abstract**—This paper deals about proposal and implementation of incremental system for multi linguistic command recognition which will be hosted in multi-agent system MASS, based on client-server architecture. Preprocessing is realized with the aid of cepstral coefficients and classification is realized by Point-Border Artmap. System allows remote parallel learning of various commands.

## I. INTRODUCTION

The goal of this paper is to propose and implement incremental system for multi linguistic command recognition in multi-agent serving system (MASS). The classification is done with our Adaptive Resonance Theory (ART) like method called Point-Border Artmap while the sound preprocessing relies on Mel-frequency cepstral coefficients (MFCC). Finally, the chosen methods implemented in form of plugins are tested in MASS with our simple Slovak commands benchmark.

MASS is considered as an approach or policy how to create families of Intelligent Systems. In this paper we describe MASS and its utilization in sound pattern recognition methods.

## II. THE STATE OF THE ART IN THE DOMAIN

If we want to solve some problems in real life with methods of artificial intelligence, we use very often recognition and classification. These concepts are very similar but there are slight differences between them. While in the process of classification the number of classification classes is known, in recognition process these classes are being created during the recognition process. The concept of classification can be defined as follows: Incorporation of objects or events into specific classes by the decision rule. The objects which are familiar enough are incorporated into the same class. Generally the classification rule has some parameters which are changeable. This change of parameters is the training of the classification tool. In the domain of linguistic command recognition, the neural networks are the common classification tools and the recurrent types of them with adaptive resonance are the best choice in many cases.

## III. SELECTED METHODS AND APPROACHES

### A. Preprocessing

Preprocessing was realized with MFCC ([2], 2006) and it has this parts:

In the first step the signal of whole word was recorded by a microphone. We used 16000Hz sampling frequency. In second part we cut the signal to micro-segments with length of 25 milliseconds. Next the following methods (Table I.) were applied on every micro-segment and reduced its dimensions from 400 to 13.

TABLE I.  
PREPROCESSING METHODS - MFCC

method	dimensions
one micro-segment	400
overlapping 10ms	0
hamming window	400
Fast Fourier Transformation	256 real, 256 img
work with only half real part	128
Mel bank filter ([3], 2004)	22
logarithm	22
Inverse Discrete Fourier Transform	22
used first 13 coefficients	13
used first 50 micro-segments from one word to classification	650

### B. Classification

Classification was realized by our neural network Point-Border Artmap. This network was derived from a well known controlled Artmap but it was very simplified.

The advantage of this network is that the cluster center is a sample from training set. With this approach it is guaranteed that every cluster is representing at least one training sample which is not so in other types of ART like networks. There are some cases when cluster is not representing any training sample and therefore it can have bad influence on classification results.

This disadvantage has original MF Artmap and also modified MF Artmap network.

Learning algorithm of Point-Border Artmap:

1. Find the closest cluster to an actual training sample with use of Euclidean distance.
2. If the winner cluster belongs to other class, go to step 4.
3. Determine Euclidean distance  $S$  from an actual sample to a winner cluster. If  $S$  is bigger than parameter  $R$ , set  $R = S$ . If there is next training sample, then take it and go to step 1, else go to step 5.
4. Create a new cluster with the center equal to the actual training sample. Set parameter  $R = 0$ . If there is next training sample, then take it and go to step 1, else go to step 5.
5. Apply the algorithm for removal of useless clusters.

Algorithm for removal of useless clusters:

1. For one network cluster (H) find other network cluster which is the closest (I).
2. If H and I belongs to the same class, find closest clusters for H (J) and I (K) which belong to a different class.
3. If J is the same cluster as K, then remove from H, I the cluster which is further.
4. If there is another not examined cluster, then take it and go to step 1.

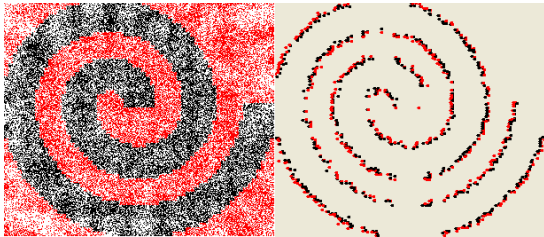


Fig. 1. Left - training set, Right - final clusters after learning algorithm was applied

#### IV. DESIGN AND IMPLEMENTATION

Everything from the previous part was implemented as a plugin for MASS. This part will describe MASS and plugin types which are used in this system.

MASS could be described as multi-agent, incremental, plugin system with client-server architecture. With plugins for object recognition it is possible to learn various objects or to recognize them in parallel manner for many clients around the world. Gained knowledge will be stored on server which will host the object recognition setup in MASS.

Fig. 2. describes the plugin system of MASS. Red plugin types are required and together they represent the smallest possible system. Green plugin types are optional. As you can see there can be more filter and processor

plugins in the system and processor plugins can be placed whether on client or server side.

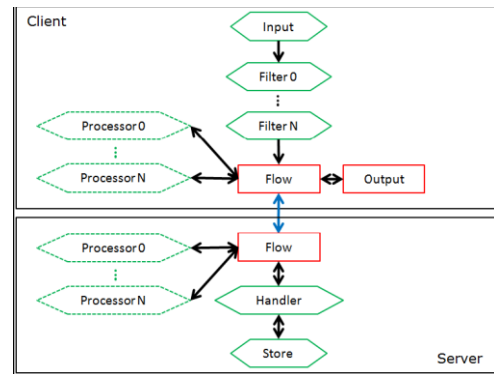


Fig. 2. MASS Plugin system

Flow plugin is special because the same plugin is placed on both sides and these sides are communicating together. Each plugin type will be now briefly described.

**Flow** is required part of the system. This plugin type is managing the client-server communication. If some user input is needed, form is included in the plugin.

**Output** is required part of the system. Its task is to provide and react on results provided by server. The reaction could be manipulation with some connected robot or device.

**Input** is optional part of the system. It provides input data from devices like webcams, microphones and sensors.

**Filter** is optional part of the system. It modifies its input, which can be from Input or other Filter plugin. The purpose of its use is similar to that of Filters known from image or audio processing.

**Processor** is optional part of the system. It can be placed whether on client or server side of the system. Processor should change the input data to data which can be used in classification, clustering or other types of Handler plugin tasks.

**Handler** is optional part of the system. This plugin should have all the functionality required for manipulation with data in database. Classification, clustering and other similar operations should be implemented in this type of plugin.

**Store** is optional part of the system. Here should be implemented everything related to data storage. This could be implemented all by authors or they can implement link to SQL or similar database system. Moreover also internet can be considered as some sort of database.

#### V. EXPERIMENTS

This section presents experiments with eight commands in Slovak language. In the first experiment, the modified MF Artmap neural network has been used for the classification. Other experiments are using Point-Border Artmap.

For improving the classification results, we are artificially creating more training data from recorded words by its shifting. We take one word sample which consists of many micro segments and we will insert the

clones of its first micro segment at the beginning together ten times. This will give us 10 new shifted samples.

#### A. Training

Experiments are using following verbal commands spoken in Slovak language (Table II). Total word count for training was 160. Individual verbal commands were spoken directly to the microphone from approximately half meter distance without disturbing environment sound. In the table are also the counts of the recorded commands from one speaker:

TABLE II.  
COUNT OF TRAINING COMMANDS

commands in Slovak language	count
	speaker 1 (man)
sadni (sit down)	20
fahni (lay down)	20
vstaň (stand up)	20
tancuj (dance)	20
kopni (kick)	20
doľava (left)	20
dozadu (go back)	20
lez (crawl)	20
total count	160

#### B. Testing

The test set consists of 80 verbal commands. Speakers were the same from the training stage. Commands were spoken directly to the microphone at approximately half meter distance without disturbing environment sound.

Counts of the commands are showed in the next table:

TABLE III.  
COUNT OF TESTING COMMANDS

commands in Slovak language	count
	speaker 1 (man)
sadni (sit down)	10
fahni (lay down)	10
vstaň (stand up)	10
tancuj (dance)	10
kopni (kick)	10
doľava (left)	10
dozadu (go back)	10
lez (crawl)	10
total count	80

#### C. Results

Original MF Artmap with parameters  $F=1$ ,  $R=0.79$ ,  $E=0.79$  reached only 52% correct classification rate. In case of modified MF Artmap, first the optimal parameter  $R = 0.61$  was experimentally determined and with this parameter modified MF Artmap reached 80% correct recognition rate. This rate was achieved after first learning cycle and more cycles had no influence on it.

The advantage of Point-Border Artmap network is that it has no parameters. Moreover we can say that this network will completely adapt to data. With this network we have reached 85% correct recognition rate. When we learned the network also with shifted samples, then the rate increased to 89%. If we used only half of training data with its shifted samples, rate was 85%.

TABLE IV.  
CONTINGENT TABLE FOR MODIFIED MF ARTMAP

actual class	predicting class								
	sadni	fahni	vstaň	tancuj	kopni	doľava	dozadu	lez	unknown
sadni	80	0	10	0	0	0	10	0	0
fahni	30	70	0	0	0	0	0	0	0
vstaň	0	0	90	0	0	0	10	0	0
tancuj	0	0	0	60	10	0	20	0	10
kopni	0	0	0	0	70	0	0	0	30
doľava	0	0	0	0	0	90	0	0	10
dozadu	0	0	0	0	0	10	90	0	0
lez	0	10	0	0	0	0	0	90	0

Final classification is 80%.

TABLE V.  
CONTINGENT TABLE FOR POINT-BORDER ARTMAP

actual class	predicting class							
	sadni	fahni	vstaň	tancuj	kopni	doľava	dozadu	lez
sadni	90	0	0	0	0	0	0	10
fahni	40	50	0	0	0	0	0	10
vstaň	10	10	80	0	0	0	0	0
tancuj	10	0	0	90	0	0	0	0
kopni	0	20	0	0	80	0	0	0
doľava	0	0	0	0	0	100	0	0
dozadu	0	0	0	0	0	0	100	0
lez	0	10	0	0	0	0	0	90

Final classification is 85%.

## VII. CONCLUSION

TABLE VI.  
CONTINGENT TABLE FOR POINT-BORDER ARTMAP WITH SHIFT, HALF TRAINING SET

actual class	predicting class							
	sadni	fahni	vstaň	tancuj	kopni	doľava	dozadu	lez
sadni	100	0	0	0	0	0	0	0
fahni	70	20	0	0	0	0	0	10
vstaň	0	0	80	20	0	0	0	0
tancuj	0	0	0	100	0	0	0	0
kopni	0	0	0	10	90	0	0	0
doľava	0	0	0	0	0	100	0	0
dozadu	10	0	0	0	0	0	90	0
lez	0	0	0	0	0	0	0	100

Final classification is 85%.

TABLE VII.  
CONTINGENT TABLE FOR POINT-BORDER ARTMAP WITH SHIFT

actual class	predicting class							
	sadni	fahni	vstaň	tancuj	kopni	doľava	dozadu	lez
sadni	100	0	0	0	0	0	0	0
fahni	30	50	0	0	0	0	0	20
vstaň	10	0	70	0	0	0	10	10
tancuj	0	0	0	100	0	0	0	0
kopni	0	0	0	0	100	0	0	0
doľava	0	0	0	0	0	100	0	0
dozadu	0	0	0	0	0	0	90	10
lez	0	0	0	0	0	0	0	100

Final classification is 89%.

## VI. CONTRIBUTION TO THE RESULTS IN THE DOMAIN

Experiments have showed quite good robustness of our neural network. There were even some interesting misclassifications. For instance the word "fahni" was few times misclassified with the word "sadni", because both words have the same ending and has the same length. This can be eliminated with a bigger training set.

Even with small amount of training data our network achieved quite good 85% correct classification rate. When we have increased the training data, rate improved to 89%.

This work is using Point-Border Artmap neural network which reaches better results in comparison with original MF Artmap and modified MF Artmap network. It was showed in experiments with almost 89% of classification accuracy. System MASS allowed us simple implementation of individual methods that were needed for the recognition in form of plugins.

Output of this work is Point-Border Artmap plugin for recognition of isolated words which can be used in MASS. Web client which can provide this functionality to public. Big pros is automatic addition of a word to the training set in case that system is not able to recognize it or if user says „zle“ after bad classified word.

In next research we will create a filter that will be able to remove disturbing environment sounds and focus only on speaker voice.

## REFERENCES

- [1] HRIC M.: Integrácia neurónových sietí typu ARTMAP s prvkami fuzzy systémov pre klasifikačné úlohy. Master thesis. Košice 2000.
- [2] PSUTKA J.: Mluvíme s počítačem česky. Academia, Praha 2006.
- [3] ČERNOCKÝ J., BURGET L.: Parametrizace řeči. FIT VUT Brno. 2003.
- [4] OLAJEC, J., JARINA R.: Použitie metódy 3TDCM pri rozpoznávaní izolovaných slov neurónovou sieťou, IEEE Vršov 2005, October 2005, Vršov, Czech Republic, ISBN 80-214-3008-7.
- [5] OLAJEC J.: Návrh rozpoznávania izolovaných čísloviek., 2004. Department of Telecommunications, EF ZU Žilina. Master thesis. 2004.

# Adaptive Proposal of Language Modification

Michal FORGÁČ

Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

Michal.Forgac@tuke.sk

**Abstract**—Software system can be viewed as an integral combination of a program and a language for this program. Program can be viewed as a sequence of statements that are aimed to produce some result. The execution is done by a platform that interprets the program's sequence of statements. The new result of a computation can be achieved by transformation of a program, an interpreter or both. Effective software evolution needs to be supported by appropriate execution environment. In this paper adaptive proposal for language modification is presented, which is based on the idea, that programming language is not an immutable artefact.

**Keywords**—Language modification, program transformation, software evolution, software language engineering.

## I. INTRODUCTION

Under modification it is possible to consider software maintenance or software evolution. These terms are often used as synonyms but I incline rather to the claim (according to e.g. [1]), that these terms do not express similar meaning because the purpose of maintenance is mostly in removal of some faults and the purpose of evolution is adaptation of a system according to the new external or internal requirements.

According to [2], software evolution is defined as a collection of all programming activities intended to generate a new version from an older and operational version. Two main types of software evolution can be named as static and dynamic evolution [7]. Static evolution consists in evolving the code of an application while it is stopped whereas dynamic evolution consists in evolving an application during its execution, without stopping it. The advantage of the former is that there is no need for state transfer or active thread to solve, whereas main disadvantage is that application is stopped and thus its services are stopped too (there is temporary unavailability). The advantage of the latter is no unavailability but there are some uncertain technical issues. Furthermore, evolution can be anticipated or unanticipated [7]. Anticipated evolution is an evolution that has been foreseen by the programmer while unanticipated evolution consists in evolution that has not been foreseen.

Software evolution includes also language evolution as an actual issue. Some projects may fail not because of bugs in programs, but because of the lack of recognition of language issues. Thus according to [9], software is composed from a program and a language. In this approach, language implementation (e.g. interpreters, compilers, and other language-dependent tools) is regarded as a metalevel of a

program.

This paper is structured as follows: section II deals with basic information about modification of interpreted functionality, concretely about program transformation, interpreter transformation and about transformation of both elements in order to produce different result. Section III presents basic information about software language engineering and section IV presents proposal of designed adaptive approach. Finally, section V concludes the paper.

## II. MODIFICATION OF INTERPRETED FUNCTIONALITY

Program  $P$  can be viewed as a sequence of statements that are aimed to produce some result  $R$ . This result  $R$  is obtained through the execution of the program  $P$ . The execution is done by a platform that interprets the program's sequence of statements. The result  $R$  of a computation depends on both a program  $P$  and interpreter  $I$ . Interpreter may be any virtual machine or in general even CPU. Different result may be obtained by changing at least one of the elements of the couple  $\langle P, I \rangle$  [8].

### A. Program Transformation

This approach is based on transformation of the program  $P_0$  from the couple  $\langle P_0, I_0 \rangle$ . The interpreter  $I_0$  is kept unchanged. A new program  $P_1$  is built from both the application aspects and the program  $P_0$ . Just aspect-oriented programming [5], [10] is based on this principle and transformation of the program  $P_0$  into the program  $P_1$  is performed using transformation process which name is weaving. Originally this approach is named as weaving through program transformation [8].

Modification of semantics through program transformation (e.g. weaving in aspect-oriented programming) may be accomplished in various ways. Modification can be realized in various times of program processing in invasively or non-invasively manner.

This modification can be static or dynamic which depends on the time when a composition tool is used. If composition is performed before compiling or it is built in compilation process, this composition is static, if it is performed after compiling, it is dynamic composition. In static composition, compile-time approaches are used. It can help in suggestion of software projects in which it is not important that system must be halted and after changes must be run again. Dynamic composition can be performed during load-time or run-time. This is very helpful for applications and services, which require non-stop running because they cannot be shut down

due to safety or financial reasons.

Modification can be classified also as invasive or non-invasive in dependence on the situation, whether it performs transformation of base code (in various forms) in order to achieve an additional functionality. Typical case of this invasive modification is modification of a source or binary code. If there is no preservation of information about origin of individual parts (e.g. in the form of meta-information) so after composition, there will not be possible to differentiate origin of individual parts of code. Typical example of non-invasive composition is run-time weaving in aspect-oriented programming, which works on the principle of interception of certain event in the base functionality and consecutive execution of additional functionality. This type of execution is realised with support of run-time environment. Invasive composition can be performed by addition of program code in well-defined form with new functionality obtained for example by repeated compilation of methods.

### B. Interpreter Transformation

This approach is based on transformation of the interpreter  $I_0$  from the couple  $\langle P_0, I_0 \rangle$ . The program  $P_0$  is kept unchanged. Originally this approach is named as weaving through interpreter transformation [8] since aspect weaving consists in applying these transformation rules to the initial interpreter  $I_0$  and so a new (or modified) interpreter  $I_1$  is created. The interpreter  $I_1$  is built from both the application aspects and the interpreter  $I_0$ .

Proposed idea of interpreter transformation is widened by transformation of various components of the execution mechanism. Thus, execution is a synonym for transformation in general, such as translation, type checking, code generation, loading, interpretation, modeling, algebraic specification, and even for informal but constructive thinking about algorithmic problems.

Foundations of adaptive language implementation were presented in [3], [4], where simple LL(1) language and its adaptive interpreter which consists of lexical analyzer, adaptive translator and evaluator were introduced. Depending on the result of interpretation, the LL(1) language was changed, and the next interpretation followed different semantics, i.e. potentially different result of the same source expression.

### C. Transformation of Both Elements

This approach is based on the experiment presented in the section IV, in which there is attempt for change the second element  $I$  from the couple  $\langle P, I \rangle$  and consecutively change the first element  $P$  from presented couple, it means from  $\langle P_0, I_0 \rangle$  to  $\langle P_0, I_1 \rangle$  and then to  $\langle P_1, I_1 \rangle$ . The main objective is to achieve this modification during run-time with utilization of the appropriate experimental run-time environment. Designed solution should support partially unanticipated evolution.

## III. SOFTWARE LANGUAGE ENGINEERING

A programming language is a medium to express computation, which is defined by its syntax and its semantics. An implementation of a programming language is the realization of its syntax and its semantics, which comprises a translator and a run-time system [13].

Nowadays there is a need for utilization of various languages, not only one single language, when there is requirement for building a software system. Languages can have textual form, graphical form or both, because they can have more than one concrete form of representation. Moreover, languages may be general purpose or domain specific.

An extensible language allows users to define new language features. These features may include new notation or operations, new or modified control structures, or even elements from different programming paradigms. According to the [12], the spectrum of language features can be divided into three classes: paraphrase, orthophrase and metaphrase. Paraphrase refers to adding new features by relying on features already present in the language (e.g. macros). Orthophrase refers to adding features not expressible in terms of available language primitives (e.g. adding an I/O system to a language that lacks one). Metaphrase refers to changing the interpretation of existing language elements, so that expressions are parsed in new ways.

Language engineering is connected to grammarware [14] and language oriented programming [15]. Grammarware comprises grammar and grammar-dependent software (e.g. parsers or XML document processors). In language oriented programming, a language is defined by its structure, editor and semantics. Its structure defines abstract syntax, its editor defines concrete syntax and its semantics define behaviour [16]. Language oriented programming introduces way of organization of development of a complex software system. This approach starts with development of formal, specific, domain-oriented, high-level programming language or more languages and at the end, development of required program. Development process is divided after phase of language development into two independent steps, one from them consists of implementation of a system with utilization of new language and the second consists in implementation of a language in some existing language, it is about building a compiler or an interpreter for a certain language.

## IV. LANGUAGE MODIFICATION APPROACH

Implemented demonstrative execution environment (Fig. 1) supports language modification approach. This experimental environment with built-in support for arithmetic expressions, is implemented with utilization of the aspect-oriented platform PROSE (Programmable extension of services) [5], [6], in which it is possible to change syntax and semantics of employed experimental language during run-time, with preservation of variables (intermediate data) and with possibility to transform input program.

### A. Implementation Background

PROSE is an infrastructure that supports dynamic adaptation by extending applications at runtime. The system performs reversible and systematic changes to running Java applications without requiring them to be shut down [6]. Implemented execution environment runs on PROSE platform, thus this environment is modifiable. Modification of syntax, semantics and program is realised with utilization of aspects, which can be inserted and withdrawn during run-time of proposed adaptive execution environment.

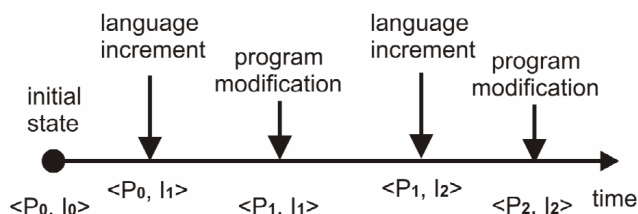


Fig. 1. Principle of presented adaptive execution environment

### B. Language Modification

Presented modification is based on the principle, that syntax and semantics of a language is inserted in various aspects. Some aspects are activated, when experimental environment is started, other aspects (especially new aspects, which represent individual language increments on the Fig. 1) are weaved during environment execution and after this time it is possible to change executed program, which can consists of new language elements.

### C. Program Modification

Although proposed solution is trivial and it is based on the principle, that when a modification of a given program is applied, executed program after modification gives altered values, in the case of extension of this solution, there is the need to consider next issues based on [11]: offline activities applied before dynamic updates, addition of a new code to the running system, deactivation of affected entities, transformation of affected entities, online verification of a new code and reactivation of halted entities. In complex extensions there may be also problems with state transfer, active threads and uncertainty.

### D. Possible Future Extensions

There are many possibilities for extension of presented adaptive execution environment. For example, the initial language should be more complex (e.g. Pascal-like programming language). Adaptive language processor should support not only changes invoked by outside changes (e.g. by user), but also changes based on inner conditions. Generation of additional parts of language processor should be also improved.

## V. CONCLUSION

This paper introduced various possibilities for modification of interpreted functionality (program semantics) in the case of the modification of a program, an interpreter or modification of both elements. Then some background of presented experimental environment for modification of both elements from presented couple  $\langle P, I \rangle$  was introduced.

Practical utilization of presented proposal in extended form can be in evolution of complex software system through incremental design of its programming language with possibility of its modification (mainly run-time modification). Another important issue is addition of new domain oriented languages during evolution of a software system.

## ACKNOWLEDGMENT

This work was supported by VEGA Grant No. 1/4073/07 Aspect-oriented Evolution of Complex Software Systems.

## REFERENCES

- [1] M. Lehman, "Laws of Software Evolution Revisited", EWSPT96, Oct. 1996, LNCS 1149, Springer Verlag, 1997, pp. 108-124.
- [2] M. Lehman and J. Ramil, "Towards a theory of software evolution - and its practical impact". Proceedings of the International Symposium on Principles of Software Evolution, Nov. 2000, Japan, pp. 2-11.
- [3] J. Kollár, J. Porubán, P. Václavík, J. Bandáková, M. Forgáč, "Functional Approach to the Adaptation of Languages instead of Software Systems", COMSIS - Computer Science and Information Systems, 4, 2, 2007, pp. 115-129, ISSN 1820-0214.
- [4] J. Kollár, J. Porubán, P. Václavík, J. Bandáková, and M. Forgáč, "Adaptive Compiler Infrastructure", Komunikačné a informačné technológie, Tatranské Zruby, October 2007, pp. 4-5, ISBN 978-80-8040-324-9.
- [5] A. Nicoara, and G. Alonso, "Dynamic AOP with PROSE", In: Proceedings of International Workshop on Adaptive and Self-Managing Enterprise Applications (ASMEA 2005) in conjunction with the 17th Conference on Advanced Information Systems Engineering (CAISE 2005), Porto, Portugal, June 2005.
- [6] A. Nicoara, "Controlled, Systematic, and Efficient Code Replacement for Running Java Programs", Dissertation ETH No. 17571, Department of Computer Science, ETH Zurich, Switzerland, December 2007.
- [7] M. Oriol, "An Approach to the Dynamic Evolution of Software Systems", Ph.D. Thesis, University of Geneva, Geneva, Switzerland, April 2004.
- [8] N. Bouraqadi, and T. Ledoux, "How to weave?", ECOOP 2001 Workshop on Advanced Separation of Concerns, June 2001.
- [9] J. M. Favre, "Languages evolve too - Changing the Software Time Scale", Eighth International Workshop on Principles of Software Evolution (IWPSE'05), 2005, pp. 33-44.
- [10] G. Kiczales, et al., "Aspect-Oriented Programming", 11th European Conf. on Object-Oriented Programming, volume 1241 of LNCS, Springer Verlag, 1997, pp. 220-242.
- [11] P. Ebraert and Y. Vandewoude, "Pitfalls in unanticipated dynamic software evolution", In the proceedings of the Workshop on Reflection, AOP and Meta-Data for Software Evolution, Glasgow Scotland, July 2005, pp 41-51.
- [12] D. Zingaró, "Moder Extensible Languages", SQRL Report 47, McMaster University, Hamilton, Ontario, Canada, October 2007
- [13] J. Malenfant, M. Jacques and F. Demers, "A Tutorial on Behavioral Reflection and its Implementation", Proceedings of Reflection 96, San Francisco, 1-20 (1996)
- [14] P. Klint, R. Lämmel, and C. Verhoeff, "Toward an engineering discipline for grammarware", In: ACM Transactions on Software Engineering and Methodology, Vol. 14, Issue 3, 2005, p. 331-380. ISSN 1049-331X.
- [15] M. Ward, "Language Oriented Programming", Software - Concepts and Tools, Vol.15, No.4, pp 147-161, 1994.
- [16] S. Dmitriev, "Language Oriented Programming: The Next Programming Paradigm.", onBoard Online Magazine, <http://www.onboard.jetbrains.com/is1/articles/04/10/lop>, November 2004.



# Theoretical introduction to nonlinear distorted OFDMA signals

Juraj GAZDA

Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

juraj.gazda@tuke.sk

**Abstract**—In this contribution, we provide a theoretical characterisation of nonlinear distortion effects in OFDMA either in the downlink as well as in the uplink scenario. Previous papers presented on this topic considered OFDMA signal as a Gaussian distributed mostly. We will show that the discussion must be strictly distinguished whether we observe the downlink or the uplink OFDMA scenario. In the uplink scenario, where usually low number of subcarriers are employed for the transmission and Gaussianity does not hold, poorer performance results with than that of downlink scenario are achieved. This effect will be properly explained by theoretical framework and proved by computer simulations.

**Keywords**—OFDMA, nonlinear distortion, Busgang theorem

## I. INTRODUCTION

Orthogonal Frequency Division Multiplexing Access (OFDMA) as a multi-user version of the popular Orthogonal Frequency Division Multiplexing (OFDM) exhibit great sensitivity to nonlinear distortion effects, caused by high power amplifiers (HPA). As long as the basic operation of OFDMA remains identical to OFDM, it is necessary to find appropriate solution that can cope with this problem in an efficient way.

The nonlinear distortion (NLD) caused by HPA at the transmitter side creates some interference both inside and outside the signal bandwidth. The in-band component determines a degradation of the system bit error rate (BER), whereas the out-of-band component affects adjacent frequency bands. In many applications, out-of-band radiation might become intolerable even when BER degradation is still acceptable.

There has been many research concerning NLD and its effects in OFDM signal. Some of the early work on this topic was done in [1] for baseband discrete multitone modulation (DMT), and in [2] for passband OFDM. All this paper assume only large number of subcarriers for transmission and operating point of HPA close to saturation. In such a case, OFDM signal as well as NLD are Gaussian distributed and as a result, OFDM signal at the output of HPA can be described by the scaled replica of the input signal plus an uncorrelated distortion term. The complex scaling term which is responsible for uniform rotation and attenuation of signal constellation can be easily compensated at the receiver side through the channel equalisation block. Distortion term has AWGN character and introduces the clouding of the signal constellation [3]. The result of this fact is that BER increase rapidly. However, discussion above is valid only if the condition of employing large number of subcarriers is fulfilled and HPA operates close to its saturation. If the operating point is far from saturation, NLD behaves as a rare-event impulsive noise and can not

be modelled as Gaussian anymore[4]. Therefore, it is strictly recommended to distinguish when investigating the downlink of OFDM, where typically large number of subcarriers are adopted and uplink, where only few subcarriers are used for transmission.

Nonlinear amplification in OFDMA transmission systems has a crucial influence on overall performance and therefore its effects must be taken into account very carefully. Regarding this, the aim of this contribution is two fold. In the first part of this paper, we provide the theoretical framework concerning NLD. The stress is put on analysis of nonlinear distortions introduced by HPA on Gaussian OFDM signal. It is worth bearing in mind, that this discussion is not valid for non-Gaussian signals, which is explained as well. In the second part of this paper, BER performance for non-Gaussian OFDMA signals is introduced and described by means of computer simulations.

## II. OFDM OVERVIEW

OFDM transmit signal is the sum of  $N$  independent sub-symbols (tones) with equal bandwidth and frequency separation  $1/T$ , where  $T$  is time duration of OFDM symbol. The  $m$ -th of encoded bits is mapped into the complex valued OFDM vector of QAM constellation points,  $\mathbf{X}^m = [X_0^m, \dots, X_{N-1}^m]$  and the continuous time representation of the single multicarrier symbol is given by [5] :

$$\mathbf{x}_{CP}^m(t) = \frac{1}{\sqrt{N}} \sum_{k=-N/2}^{k=N/2-1} X_k^m e^{j2\pi kt/T} \quad (1)$$

where  $m$  is a symbol index and  $X_k^m$  is the QAM value of  $k$ -th subsymbol or tone. The periodic extension of the symbol over the interval  $[-T_{CP}, 0]$  is the cyclic prefix (CP) which simplifies the equaliser design in the presence of multipath fading.

In practise, OFDM signals are generated using Inverse Discrete Fourier Transform (IDFT). The resulting  $T/N$ -spaced discrete time vector  $\mathbf{x}^m = [x_0^m, \dots, x_{N-1}^m] = IDFT(\mathbf{X}^m)$  is given by:

$$\mathbf{x}^m = \frac{1}{\sqrt{N}} \sum_{k=0}^{k=N-1} X_k^m e^{j2\pi ktn/N} \quad (2)$$

In this paper, the discrete time indexing  $[n]$  denotes Nyquist rate samples. In order to avoid aliasing and the out-of-band radiation into to data bearing tones, the oversampling of the original signal  $x[n]$  may be needed [6]. We will introduce the notation  $x[n/L]$  to denote oversampling by factor  $L$ . In the simulation presented in this paper, oversampling factor of 4 has been applied.

### III. CHARACTERISATION OF DECISION VARIABLES AT THE DEMODULATOR STAGE

In order to be able to describe statistical behaviour of nonlinear distortion, we first describe the signal at the input of HPA, and in the next we exploit the signal at the output of HPA. Finally, we investigate the signal in the presence of AWGN at the input of the baseband modulator.

#### A. Signal at the input of HPA

In general, OFDM signal is a sum of many independent subcarriers, that are modulated by baseband QAM symbols. Without loss of generality, we can assume, that baseband QAM symbols are identically distributed random variables. According to the central limit theorem, it is reasonable to assume that OFDM signal is Gaussian distributed if large number of subcarriers is employed for transmission. As a result, OFDM signal envelope follows Rayleigh distribution as [7]:

$$f_X(x) = \frac{2x}{\sigma^2} e^{-\frac{x^2}{\sigma^2}} \quad (3)$$

Since the envelope of OFDM signal follows Rayleigh distribution, it is apparent, that the envelope significantly fluctuates in time. It is worth to realise, that the signals with different amplitudes will be affected by the nonlinearity differently, following the probabilistic distribution of the OFDM signal. This fact has crucial influence on the overall performance of OFDM transmission system [8].

#### B. Signal at the output of HPA

Now, we focus our attention on the nonlinear block of HPA. Following mathematical derivation are derived by the application of Busgang theorem (*Appendix A*) [9]. The output signal  $y(t)$  of memoryless HPA can be described as follows:

$$y(t) = \alpha x(t) + d(t) \quad (4)$$

where  $\alpha = R_{xy}(\tau_1)/R_{xx}(\tau_2)$  is complex gain,  $R_{xy}$  denotes crosscorrelation function of input and output signal and  $R_{xx}$  denotes autocorrelation function of the input signal. Note that  $\alpha$  is independent on particular symbols realisation  $[X_0, \dots, X_{N-1}]$  and remains constant during entire transmitting process [3]. The scaling factor  $\alpha$  is responsible for the attenuation and rotation of the constellation, which can be easily compensated at the receiver by introducing correction factor  $\alpha/|\alpha|^2$ . However, the distortion term  $d(t)$ , which is responsible for both the clouding and the out-of-band radiation can not be compensated. It is easy to show that  $d(t)$  is uncorrelated with the input signal  $x(t)$ :

$$\begin{aligned} R_{xd} &= E[x(t) * (y(t + \tau) - \alpha x(t + \tau))] \\ &= R_{xy}(\tau) - \alpha R_{xx}(\tau) = 0 \end{aligned} \quad (5)$$

#### C. Output vector of the FFT block

This part of this paper determines the BER performance in nonlinear AWGN environment. The signal at the output of FFT is given by [3]:

$$R_k = \frac{1}{\sqrt{N}} \sum_{n=0}^{LN-1} r_n e^{-j2\pi kn/LN}, k = 0, \dots, N-1 \quad (6)$$

where  $r_n$  is the input signal of FFT. In general  $R_k$  consists of the useful part of the signal, noise components and nonlinear distortion. This fact can be expressed as [3],[10]:

$$R_k = \alpha S_k + D_k + W_k \quad (7)$$

where  $D_k$  is noise component from nonlinear distortion and  $W_k$  is AWGN component.

Let us analyse nonlinear distortion noise in in further discussion. Nonlinear distortion at the output of FFT can be expressed as [3],[10]:

$$D_k = \frac{1}{\sqrt{N}} \sum_{n=0}^{LN-1} d_n e^{-j2\pi kn/LN} \quad (8)$$

Note that nonlinear distortion term  $D_k$  is composed as the sum of  $N$  identically distributed random variables, so we can assume it to be complex Gaussian distributed. In [10] it was shown that even the certain parts of the sum are correlated, after all,  $D_k$  might be considered as a complex Gaussian.

As it is clear from above mentioned discussion, both  $W_k$  and  $D_k$  follows Gaussian distribution which lead us to the formulation of the signal to noise-plus-distortion ratio (SNDR) as [3]:

$$SNDR = \frac{|\alpha|^2 E[|S^2|]}{\sigma_W^2 + \sigma_D^2} \quad (9)$$

where  $\sigma_W^2$  and  $\sigma_D^2$  is the variance of nonlinear distortion term and variance of Gaussian noise, respectively.

### IV. INPUT BACK-OFF AND OPERATING POINT OF HPA

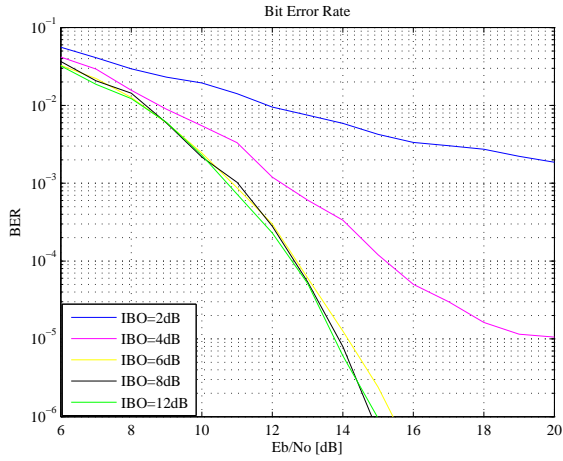
The nonlinear distortion of HPA depends strongly on the input back-off (*IBO*), which is defined as [11]:

$$IBO = \frac{P_{in}}{P_{sat}} \quad (10)$$

where  $P_{sat} = A_{sat}^2$  represents the saturation power and  $P_{in} = E|x(t)|^2$  the mean power of the input signal  $x(t)$ . Small values of the *IBO* causes the amplifier operation point to be near the saturation. In this case a good HPA efficiency is achieved, but as a consequence the HPA output signal will be highly distorted. Keep in mind that if high *IBO* is applied at the transmitter side, HPA operates far from saturation point and as a result nonlinear distortion behaves as a rare-event impulsive noise and follows no more Gaussian distribution [4].

In order to illustrate the above mentioned facts, the appropriate computer simulations have been arranged. Within this simulations, OFDM signals with  $N = 256$  subcarriers, operating at Rapp model of HPA, is presented. Fig.1 represent BER of OFDMA as a function of  $E_b/N_0$ . On one hand the simulation results do match the analytical results presented in [3] for low values of *IBO* almost perfectly (*IBO* = 2dB, 4dB), on the other hand, with increasing value of *IBO*, the differences between analytical results and results obtained by computer simulation increase. In these cases, the HPA is more linear, less distortion is introduced and therefore distortion term introduced by HPA become less Gaussian distributed.

In Fig.2, power density spectrum of OFDM signal for different *IBO* parameters is shown. The simulation results confirm the theoretical expectations, the higher *IBO*, the lower out-of-band radiation is presented. However during the design process, one has to find trade-off between BER, power efficiency and out-of-band radiation [12].


 Fig. 1. BER vs.  $E_b/N_0$ , different values of  $IBO$ 

## V. PERFORMANCE OF NON-GAUSSIAN OFDM SIGNAL

The mathematical formulas presented above have been derived under condition of Gaussianity of OFDM signal. However, when low number of subcarriers is adopted, Gaussianity does not hold. It was shown by means of normalised kurtosis in [3], that OFDM signal with low number of subcarriers follows sub-Gaussian distribution. The signal at the output of HPA is given in this case as [4]:

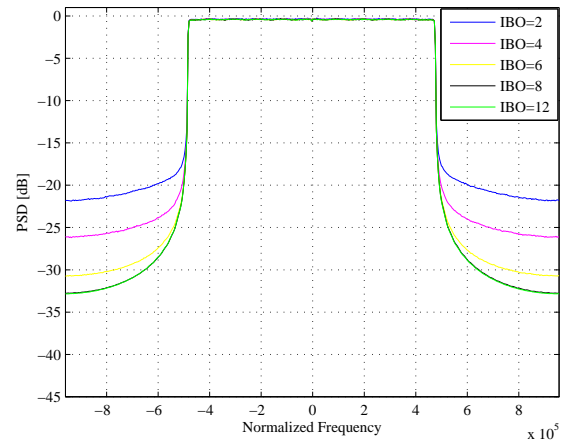
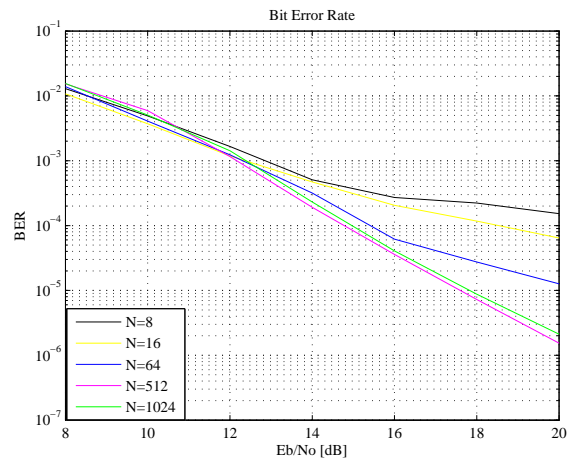
$$y^i(t) = \alpha^i s^i(t) + d^i(t) \quad (11)$$

where  $\alpha^i$  is a complex gain factor that depends on particular realisation  $[X_0^i, \dots, X_{N-1}^i]$  and  $d^i(t)$  is the distortion term. In such case, the complex gain  $\alpha^i$  is different for different OFDM symbols.

As it was described in [3],[4], sub-Gaussian signals exhibit much worse BER performance results in AWGN channel, than signals that follows Gaussian distribution. This fact is illustrated by Fig.3, where BER vs  $E_b/N_0$  for different number of subcarriers is illustrated. As it is clear from this figure, the lower number of subcarriers is employed for transmission, the higher BER is achieved. This fact is crucial mainly in the uplink of OFDMA, where only few subcarriers are usually adopted. In order to mitigate this problem, 3GPP introduced single carrier frequency division multiplex (SC-FDMA) in an upcoming 4G wireless communication system uplink instead of OFDMA [13]. SC-FDMA is very robust against NLD and what is very important, NLD remains constant with changing number of subcarriers [4],[13]. However, it is important to note here, that in the next evolution of LTE, called LTE-Advanced, there might be considered even OFDMA in the uplink transmission. Therefore OFDMA is still very promising solution how to offer large flexibility and reliability in the next evolutions of modern wireless communication systems.

## VI. CONCLUSION

Nonlinear amplification might be a very difficult challenge, unless the theoretical aspects of nonlinear distortion are not explained properly. This fact motivates the authors to provide the theoretical discussion of the nonlinear effects in OFDMA. In the first part of the paper theoretical approach to the evaluation of nonlinear distortion effects of OFDMA signal in AWGN channel is presented. To build this procedure, Busgang theorem has been adopted


 Fig. 2. Power density spectrum, different values of  $IBO$ 

 Fig. 3. BER vs.  $E_b/N_0$ , different number of subcarriers

in mathematical formulation. However, it was shown, that for low number of subcarriers, Busgang theorem is not valid and OFDM signal becomes following sub-Gaussian distribution. In a second part of this paper, we observed, that BER degradation due to NLD is somehow related to the number of subcarrier. The lower number of subcarriers is, the higher BER degradation occur. From this point of view, OFDMA uplink transmission is much more vulnerable to nonlinear amplification than the downlink, where Busgang theorem holds. Therefore we highly recommend to use in this case particular available methods that mitigate this harmful nonlinear effects either on transmitter, or receiver side.

## APPENDIX BUSGANG THEOREM

For two Gaussian signals  $x_1(t)$  and  $x_2(t)$ , the cross-correlation function taken after one of them (e.g.  $x_2(t)$ ) has undergone nonlinear amplitude distortion ( $R_{x_1 y_2}$ ) is identical except for a factor of proportionality  $\alpha$ , to the cross-correlation function taken before the distortion ( $R_{x_1 x_2}$ ):

$$R_{x_1 y_2}(\tau) = \alpha R_{x_1 x_2}(\tau) \quad (12)$$

Notice that if  $x_1(t) = x_2(t)$  then it follows that the cross-correlation between input and output signals of the nonlinearity is identical, except for a factor of proportionality

$\alpha$ , to the autocorrelation of the input signal, that is

$$R_{xy}(\tau) = \alpha R_{xx}(\tau) \quad (13)$$

#### ACKNOWLEDGMENT

This work has been funded by VEGA 1/4088/07 Rekonfigurovateľné platformy pre širokopásmové bezdrôtové telekomunikačné siete and COST 297:High Altitude Platforms for Communications and other Services.

#### REFERENCES

- [1] Mestdagh, D.J.G., Spruyt, P., Biran, B.: Analysis o clipping effect in DMT-based ADSL systems, in *Proc. IEEE International Conference Communication*, 1994
- [2] O'Neill, R., Lopes, L.B: Performance of amplitude limited multitone signals, in *Proc. IEEE Vehicular Technology Conference*, pp.1675-1679, Jun 1994
- [3] Deumal, M., Behravan, A.: Nonlinear distortion effects and signal metrics in OFDMA systems, *submitted for publication in IEEE Transactions on Communications*
- [4] Deumal, M.: Multicarrier communications systems with low sensitivity to nonlinear amplification, *Phd thesis*, Ramon Llull university, Barcelona, 2008
- [5] Bingham, J.A.C. Multicarrier modulation for data transmission, *IEEE communication magazine*, pp.5-14, vol.28, May 1990
- [6] Banelli, P., Cacopardi, S.: Theoretical analysis and performance of OFDM signals in nonlinear AWGN channels, *IEEE Transactions on Communications*, vol.48, March 2000
- [7] Proakis, J.G.: *Digital communicatios*, 4th ed. McGraw-Hill, 2000
- [8] Pavelka, P., Krajnak, J., Galajda, P., Kocur, D.: Analysis of non-linear distortions in MC-CDMA systems, *Acta Electrotechnica et Informatica*, vol.7, No.4, 2007
- [9] Bussgang, J.: Crosscorrelation function of amplitude-distorted gaussian signals, Research laboratory of electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts, 1952
- [10] Dardari, D., Tralli V., Vaccari, A.: A theoretical characterization of nonlinear distortion effects in OFDM systems, *IEEE Transactions on Communications*, vol.48, pp.1755-1763, Oct. 2000
- [11] Behravan, A., Eriksson, T.: Some statistical properties of multicarrier signals and related measures, in *Proc. IEEE Vehicular Technilogy Conference Spring*, vol.4 May 2006
- [12] Fazel, K., Kaiser, S.: *Multi-carrier and spread spectrum systems*, John Wiley & Sons, 2003
- [13] Fazel, K., Kaiser, S.: *Multi-Carrier and Spread Spectrum Systems, From OFDM and MC-CDMA to LTE and WiMAX*, John Wiley & Sons, 2008

# Learning of Fuzzy Rules with Generalization for Dichotomic Classification

<sup>1</sup>Daniel HLÁDEK

<sup>1</sup>Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

<sup>1</sup>daniel.hladek@tuke.sk

**Abstract**—We propose a novel fuzzy rule learning approach for quick generating of sub-optimal rulebase suitable for tasks with lower dimension of the state space. This paper includes a series of experiments for dichotomic classification for spiral benchmark data, that evaluate the process of learning and ability to generalize knowledge extracted from data.

**Keywords**—reinforcement learning, fuzzy sarsa, generalization, learning classifier system

## I. INTRODUCTION

This paper proposes a method to learn a fuzzy rulebase in the Mamdani-type inference engine for a task where result can be evaluated by a value that expresses quality of the current output of the system.

The number of rules in our approach can decrease during the learning process, because of the rule-generalization heuristics that is incorporated in our system. This heuristics allows to join two rules that looks similar, to new rule that covers the antecedent part of the both old rules.

This paper is organized as follows. First, the most important approaches in the field of the learning classifier systems are presented. Next part proposes the learning classifier system with rule generalizing heuristics. The experiments part evaluates the proposed system. Finally, we conclude with the assessment of the system and proposing future research focus.

## II. STATE OF THE ART

Learning classifier systems are a method for evolution of the rulebase in a reinforcement learning task. According to the author of the concept J. Holland in the [1], we can characterize **learning classifier systems** (LCS) as: “[Learning] Classifier systems are a kind of rule-based system with general mechanisms for processing rules in parallel, for adaptive generation of new rules, and for testing the effectiveness of existing rules.”

Term **classifier** means a rule with antecedent and consequent part and in addition contains parameters that evaluate performance of that rule in situations, where the antecedent part of the rule applies.

Theory of learning classifier systems uses **reinforcement learning** algorithms for adapting parameters of the classifier set. Output of the system is characterized by immediate reward, given by the system. This reward is then distributed between classifiers that contributed to the current state in the past using one of the reinforcement learning algorithms, such as Q-learning [2] or Sarsa [3]. The result of continuous evaluation of the classifiers is expected reward for each classifier in the system, gained after applying the classifier.

The procedure of the Sarsa algorithm for updating expected reward ( $Q(s, a)$  value) of action  $a$  in state  $s$  is described as algorithm 1.

---

### Algorithm 1 Sarsa algorithm

---

```

1:  $s$  is start state
2: choose  $a$  according to the policy
3: while state is terminal do
4:   Take action  $a$ 
5:   Observe immediate reward  $r$ 
6:   Observe state  $s'$ 
7:   Choose action  $a'$  from state  $s'$ 
8:   Update  $Q(s, a)$ :  $Q(s, a) \leftarrow Q(s, a) + \alpha(r + \gamma Q(s', a') - Q(s, a))$ .
9:    $a \leftarrow a', s \leftarrow s'$ 
10: end while

```

---

Existing LCS approaches can be divided into three groups according to the method of testing of effectiveness of rules and parameters of the classifier.

Example of the **strength based approach** is ELF (Evolutionary learning of fuzzy rules) algorithm. It has been used for evolution of robotic behavior in the work of Andrea Bonarini [4] and its modification in [5]. In the strength based approach a rule fitness in the evolutionary process is based on the amount of expected reward. This parameter is incrementally computed using evaluation of the classifier performance.

The **accuracy based approach** and one of the most published themes in the area of LCS is algorithm called XCS. First version was released by Wislon in [6] and its fuzzy extension in [7]. Accuracy based approach takes accuracy of the reward prediction as a rule fitness, rather than prediction itself. The most accurate classifier has the biggest probability of surviving in the evolution.

The **anticipation based approach** is an algorithm by Butz [8] called ACS. Every classifier in the anticipatory classifier system contains state of the that is expected after execution of the classifier action. The fitness of the classifier is based on accuracy of this expectation and accuracy of the reward prediction.

## III. PROPOSED APPROACH

### A. Classifier Form

The common problem of the learning classifier systems is the generalization ability. This ability allows to cover more situations by just one classifier. Common way of solving this

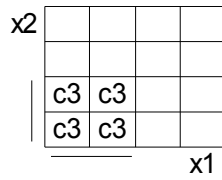


Fig. 1. Rule with disjunctive normal form

problem is by using # symbol, that covers all values of the input space in one dimension. We find this way not precise enough.

Our approach proposes a way that allows to omit the # symbol by using a full disjunctive normal form of the rule in the LCS. This means that OR conjunction is allowed for the classifier.

In the case of the input space as a 4x4 grid, we can then have rule that looks like (on the figure 1):

```
IF x1=(a0 OR a1) AND x2=(b0 OR b1)
THEN y=c3 CREDIT q=-0.34
```

This situation can be encoded using string as:

```
1100 1100 | 3 0.34
```

The left part describes antecedent part of the rule using binary string, where 1 means presence of the linguistic value in the disjunctive normal form expression. Right part marks the consequent (e.g. result class) and prediction of gained reward marked q.

Classifiers in the LCS are grouped into **subpopulations** according to their antecedent part. Classifiers with the same antecedent part belong to the same subpopulation. This division into subpopulations allows easier control of the classifier set consistency (each part of the input space can be covered by one subpopulation) and easier distribution of the immediate reward (credit is received by just one classifier in the subpopulation).

### B. The Algorithm

The basic scheme for our approach comes out of the work of A. Bonarini [9]. It is purely strength-based approach, which means that every classifier also contains a single expected reward parameter. The original Bonarini's work used Q-learning like method for the classifier evaluation, but we have used modified Sarsa algorithm (described bellow). The basic algorithm is written in pseudocode as algorithm 2.

In the beginning, classifier set is generated, such that all possible state of the input space is covered by one subpopulation. The subpopulation is filled with all possible classifiers.

In the first step of one round of the algorithm, the state is observed and match set is constructed. Match set contains all subpopulations that match the current state. Since we use Mamdani's fuzzy inference, each subpopulation can match the situation to a certain level. Then each subpopulation in the match set is observed and classifier with the best q value in the subpopulation is inserted to the action set.

Classifiers in the action set are taken and using rules of the fuzzy inference the final action is constructed. This action is sent for execution, stored and the immediate reward is observed. This reward is distributed and q values of the classifiers in the action set from the last round are updated. Match and action set is emptied and new round can begin.

---

### Algorithm 2 The Learning Algorithm

---

**Require:** classifier set C

```
1: action set  $A_{old} = \emptyset$ 
2: while end condition do
3:   observe the state  $s_t$ .
4:   calculate match set  $M$ 
5:    $A_{old} = A, A = \emptyset$ 
6:   for all subpop in  $M$  do
7:     choose an action (classifier c) from subpop
8:     insert classifier into A
9:   end for
10:  execute action in A, but don't wait for the result
11:  observe reward  $r_t$ 
12:  distribute reward  $r_t$  to  $A_{old}$  and update classifiers
13: end while
```

---

### C. Credit Distribution

First part of the classifier update procedure is the credit distribution. The immediate reward is taken and distributed among classifiers in the action set from last round. The whole process of the credit distribution and classifier update is written as the algorithm 3.

The q value update formula is based on the Sarsa algorithm (see line 8 of the algorithm 1). However, this formula needed to be updated for the conditions of the fuzzy inference. We have used Fuzzy Sarsa [10], [11] approach:

$$\Delta q_m = \alpha * \{r_t + \gamma \sum_{n=1}^N act(n, s_{t+1})q_n - \sum_{m=1}^M act(m, s_t)q_m\} act(m, s_t) \quad (1)$$

where  $act()$  is normalization coefficient based on activation level of the classifier in the action set.

$$act(m, s_t) = \frac{\mu_m(s_t)}{\sum_{m=1}^M \mu_m(s_t)} \quad (2)$$

$\alpha$  is the learning coefficient,  $\gamma$  is discount rate. This coefficient expresses how much the current state of the system depends on the previous state. In the multi-step problems, such as path search is close to 1, in the classification task is zero.

N is number of classifiers in the current action set A, M is number of classifiers in the last round action set  $A_{old}$  and  $q_m$  is their q-value.

The  $\Delta q_m$  value is added to every classifier in the action set  $A_{old}$ .

### D. Generalization Operator

If the subpopulation has been enough tested, then the generalization routine can start (lines 3-11 of the algorithm ??). In this process is the classifier set searched for similar subpopulation that is also enough tested. If such a classifier is found, then these two subpopulations can be joined together to form a new subpopulation and then dropped.

The question is when we can tell that the subpopulations are similar. If two subpopulations are not similar enough, the resulting subpopulation could not be a valid generalization of the proceeding subpopulations.

**The condition of similarity** is divided into two parts. First, the antecedent part must be similar. This means that area

**Algorithm 3** Classifier Update Procedure

---

**Require:** action set  $A_{old}$ , action set  $A$ , classifier set  $C$ , immediate reward  $r_t$ ,

- 1: **for all**  $c_i \in A_{old}$  **do**
- 2:    $q_i \leftarrow q_i + \Delta q_i$
- 3:   **if**  $c_i$  is enough tested and not large enough **then**
- 4:     **for all**  $c_j \in C$  **do**
- 5:      **if**  $c_i$  is similar to  $c_j$  **then**
- 6:        $subpop_i \leftarrow \text{join } subpop_i \text{ and } subpop_j$
- 7:       drop  $subpop_j$
- 8:       continue
- 9:      **end if**
- 10:     **end for**
- 11:    **end if**
- 12: **end for**

---

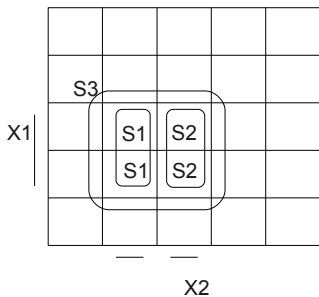


Fig. 2. Subpopulation Join Procedure

covered by two input classifiers must be consistent - cover a continuous area, such as is depicted on figure 1. This can be achieved only if the input classifiers are near each other.

The second condition is, that the best classifiers in both input subpopulations needs to have the same consequent part.

If both conditions are met, the subpopulations can be joined together and they form a new subpopulation. This situations is shown on figure 2. Input area covered by subpopulation 1 (S1) is near the area of the subpopulation 2 (S2). Best classifiers in the both subpopulations have the same consequent part. Then these can be joined together and new subpopulation S3 is created. This new subpopulation again contains classifiers for all possible consequents.

#### IV. EXPERIMENTS

We have tested the approach on the classification of double-spiral data. Train data take form of triplet  $(x_1, x_2, c)$ , where  $x_1, x_2$  is the position of the point and  $c$  is class of this point.

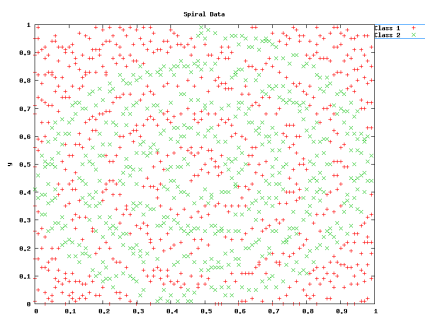


Fig. 3. Spiral Train Data

Train set is on the figure 3. There are two possible classes for the point, class 1 is marked as a red +, class 2 is marked as green x on the figures. The train set contains 1250 samples.

The goal in the learning phase is to assign correct class for every point that is given to the input. The system receives feedback about its answer in the form of immediate reward, that is defined as a function that returns:

$$r_t = \begin{cases} 0 & \text{if result is correct} \\ -1 & \text{if result is incorrect} \end{cases} \quad (3)$$

If the system returns value from interval  $(0, 1)$ . If the output value is lesser than 0.5 it is counted as class 1, class 2 is otherwise.

In the beginning of the experiment, the input space is covered by the grid of 35 x 35 cells. Every cell in the grid corresponds to one subpopulation (similar situation to the figure 1, but one subpopulation covers just one cell). Each subpopulation contains two classifiers, one for each possible class. This means, that the learning classifier system initially contains 2450 classifiers in 1225 subpopulations.

From the view of the fuzzy logic, the inference system contains two input linguistic variables and one output linguistic variable. Linguistic value of the input variable is one triangle-shaped fuzzy set, where the linguistic variable is created by the normal fuzzy partition of the input attribute ( $x_1$  or  $x_2$ ).

Then a classifier takes form:

IF  $x_1 = \text{part2}$  AND  $x_2 = \text{part20}$  THEN  $\text{class} = \text{c1}$

In the experiment, we have set learning parameter  $\alpha$  to the value 1.5 and discount parameter  $\gamma$  to zero. (discount parameter is zero, because in we solve one-step problem, current state of the system - input - does not depend on the previous state) Credit distribution function is then simplified to:

$$\Delta q_m = \alpha * \{r_t - \sum_{m=1}^M \text{act}(m, s_t) q_m\} \text{act}(m, s_t) \quad (4)$$

In this experiment, classifier is enough tested when sum of its activations reach value 30.

To evaluate the generalization ability of the system we have executed two experiments on the same data - experiment 1 with the generalization operator enabled, and experiment 2 with disabled.

The learning process of the experiment 1 is on the figure 4 and experiment 2 on the figure 5. The upper + line marks number of subpopulations on the learning classifier system. The  $\diamond$  line expresses performance of the system as a sum of the reward gained through one sweep of the train set. The system receives -1 for each incorrectly classified sample.

We can see, that the performance of the system improves rapidly in the both experiments.

After 30 sweeps of the train set we have counted number of classifiers and calculated mean square error

$$E = \frac{1}{2} (c^a - c^e)^2 \quad (5)$$

where  $c_a$  is the actual result from the learning classifier system, and  $c_e$  is the expected class from the test set.

- Results of experiment 1 : 857 subpopulations,  $E = 0.1723$
- Results of experiment 2 : 1225 subpopulations  $E = 0.156$

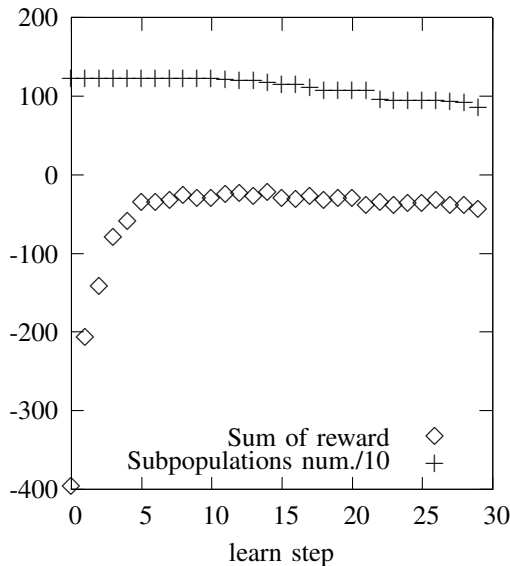


Fig. 4. Learning Experiment 1

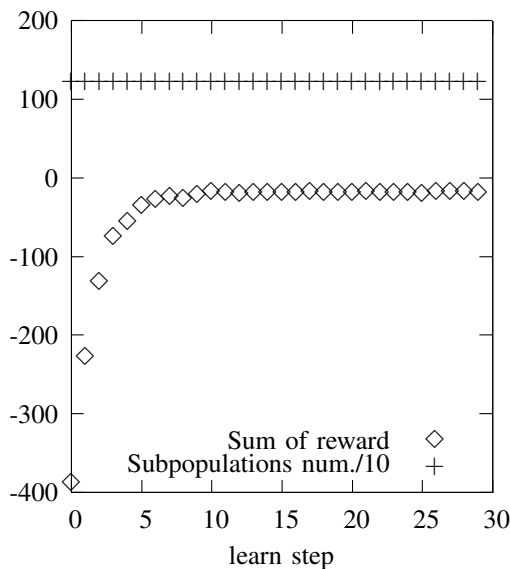


Fig. 5. Learning Experiment 2

We can see, that the using rule generalization has decreased size of the rulebase. The generalization operator did not increase classification error very much.

Part of the obtained rulebase form the experiment 1 (best classifier form each subpopulation is selected):

```

if x1 is 19 and x2 is 11 then y is class1
if x1 is 17 18 19 20 and x2 is 12
                                then y is class1
if x1 is 19 and x2 is 13 14
                                then y is class2
if x1 is 19 and x2 is 15 then y is class1
    
```

## V. CONCLUSION

Experiments has proven usability of the generalization operator in the learning of the classifier rulebase. The generalization can cover multiple states of the system by one classifier and reduce the number of necessary rules for the system.

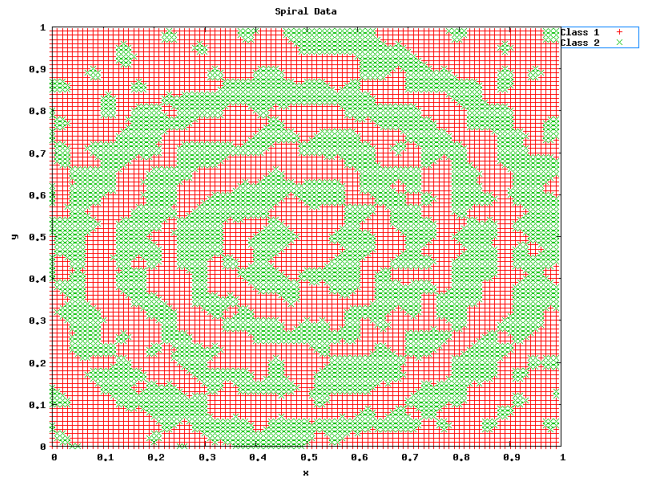


Fig. 6. Result of the Spiral Experiment 1

Main application area of the learning classifier systems lie in the multi-step problems such as path search or planning. In further research we want to focus on the application of the algorithm in the area of robotics. We would like to explore the abilities of the algorithm in the path searching.

## REFERENCES

- [1] J. H. Holland, K. J. Holyoak, R. E. Nisbett, P. R. Thagard, and S. W. Smoliar, "Induction: Processes of inference, learning, and discovery," *IEEE Expert*, vol. 2, no. 3, pp. 92–93, Sept. 1987.
- [2] C. Watkins, "Learning from Delayed Rewards," Ph.D. dissertation, Cambridge University, 1989.
- [3] G. Rummery and M. Niranjan, *On-line Q-learning using connectionist systems*. Univ. of Cambridge, Department of Engineering, 1994.
- [4] A. Bonarini and F. Basso, "Learning to compose fuzzy behaviors for autonomous agents," *Int. Journal of Approximate Reasoning*, vol. 17, no. 4, pp. 409–432, 1997. [Online]. Available: [citeseer.ist.psu.edu/bonarini97learning.html](http://citeseer.ist.psu.edu/bonarini97learning.html)
- [5] M. Chronis, G. Keller, and J. Skubic, "Learning fuzzy rules by evolution for mobile agent control," in *Computational Intelligence in Robotics and Automation, 1999. CIRA '99. Proceedings. 1999 IEEE International Symposium on*, 1999, pp. 70–76.
- [6] S. W. Wilson, "Generalization in the XCS classifier system," in *Genetic Programming 1998: Proceedings of the Third Annual Conference*, J. R. Koza, W. Banzhaf, K. Chellapilla, K. Deb, M. Dorigo, D. B. Fogel, M. H. Garzon, D. E. Goldberg, H. Iba, and R. Riolo, Eds. University of Wisconsin, Madison, Wisconsin, USA: Morgan Kaufmann, 22-25 1998, pp. 665–674. [Online]. Available: [citeseer.ist.psu.edu/wilson98generalization.html](http://citeseer.ist.psu.edu/wilson98generalization.html)
- [7] J. Casillas, B. Carse, and L. Bull, "Fuzzy-xcs: A michigan genetic fuzzy system," *IEEE Trans Fuzzy Syst*, vol. 15, no. 4, pp. 536–550, 2007. [Online]. Available: <http://www.scopus.com/scopus/inward/record.url?eid=2-s2.0-34548246008&partnerID=40&rel=R7.0.0>
- [8] A. M. V. Butz, B. D. E. Goldberg, and C. W. Stolzmann, "The anticipatory classifier system and genetic generalization," *Natural Computing*, vol. 1, no. 4, pp. 427–467, Dec. 2002. [Online]. Available: <http://dx.doi.org/10.1023/A:1021330114221>
- [9] A. Bonarini, "Delayed Reinforcement, Fuzzy Q-Learning and Fuzzy Logic Controllers," in *Genetic Algorithms and Soft Computing, (Studies in Fuzziness, 8)*, F. Herrera and J. L. Verdegay, Eds. Berlin, D: Physica-Verlag, 1996, pp. 447–466.
- [10] L. Tokarchuk, J. Bigham, and L. Cuthbert, "Fuzzy Sarsa: An approach to fuzzifying Sarsa Learning," in *Proceedings of the International Conference on Computational Intelligence for Modeling, Control and Automation*, 2004.
- [11] T. Theodoridis and H. Hu, "The Fuzzy Sarsa( $\lambda$ ) Learning Approach Applied to a Strategic Route Learning Robot Behaviour," *Proceedings of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems October 9 - 15, 2006, Beijing, China*, pp. 1767–1772, 2006.



# Production lines modeling with the use of Coloured Petri Nets

<sup>1</sup>Juraj CHOVAŇÁK

<sup>1</sup>Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

<sup>1</sup>[juraj.chovanak@tuke.sk](mailto:juraj.chovanak@tuke.sk)

**Abstract**—As a graphical and mathematical tools Petri nets, provide a uniform environment for modeling, analysis, and design of discrete event systems. This paper deals with coloured Petri nets (CPN). Their typical examples of application areas are communication protocols, distributed systems, automated production systems and workflow analysis. The main objective of this paper was to create a production line, build from parts of a school manipulator workstation model set and to create a model of this system with the use of coloured Petri nets.

**Keywords**—Distributed Control System, Coloured Petri Net, Robotics Workstation Model

## I. INTRODUCTION

The accelerated expansion and integration of communications, computing, and control over the past few years has inspired researchers and practitioners from a variety of disciplines to become interested in the field of Distributed Control Systems (DCS). DCS refers to a control system usually a manufacturing system, process or any kind of dynamic system, in which the controller elements are not central in location but are distributed throughout the system with each component sub-system controlled by one or more controllers [1]. The latest industrial automation systems use highly distributed architectures where a number of digital modules are interconnected by communication networks for data acquisition and lower level control functions [2]. The growth in the complexity of such systems creates numerous problems for their developers [3]. In the planning stage, one is confronted with increased capabilities of these systems due to the unique combination of hardware and software, which operate under a large number of constraints arising from the limited system resources. In view of the capital intensive and complex nature of modern industrial systems, the design and operation of these systems require modeling and analysis in order to select the optimal design alternative, and operational policy. It is well-known that flaws in the modeling process can substantially contribute to the development time and cost. The operational efficiency may be affected as well. Therefore special attention should be paid to the correctness of the models that are used at all planning levels. Petri nets, as graphical and mathematical tools, provide a uniform environment for modeling, formal analysis, and design of discrete event systems. The major advantage of using Petri net models is that the same model is used for the analysis of behavioral properties and performance evaluation. The process of creating a description

and performing the analysis gives to the developer a dramatically improved understanding of the modeled system.

## II. PETRI NETS BASIC DESCRIPTION

A Petri net may be identified as a particular kind of bipartite directed graph populated by three types of objects. These objects are places, transitions and directed arcs connecting places to transitions and transition to places. Places, represented by circles describe the states of the system. Transitions, represented as boxes or bars, describe the actions and the arc expressions describe how the state of the Petri net changes when the transitions occur. In order to study dynamic behavior of the modeled system, in terms of its states and their changes, each place may potentially hold either none or positive number of markers, called tokens. Each of these tokens carries a data value, for instance, for a place representing the availability of resources, the number of tokens in this place indicates the number of available resources. The distribution of tokens on places, called the Petri net marking, defines at any given time instance the current state of the modeled system.

CPN [4] have got their name because they allow the use of tokens that carry data values and can hence be distinguished from each other – in contrast to the tokens of low-level Petri nets, which by convention are drawn as black, “uncoloured” dots. In the beginning, only small, unstructured sets of colours were used, but later it was realised that it was possible to generalize the theory and the tools, in such a way that arbitrarily complex data types can be used as colour sets. This is very important for Manufacturing Systems that have to deal with several products at the same time. CPN can also be extended with a time concept. This means that it is possible to use the same modeling language for both, the specification and validation of functional properties such as boundness, liveness and reversibility, and also to analyze the performance of the system, such as average waiting times and throughput. More about CPN can be found in [5], [6].

## III. EXPERIMENTAL WORKSTATION MODEL

This model represents a robotics workstation. It is situated at Department of Cybernetics and Artificial Intelligence. The workstation consists of two manipulators, two conveyors and one rotary feeder (Figure 1). These parts can be assembled into various configurations and so simulate several robotics workstations. The manipulator arm is an arm with five degrees of freedom from which three degrees are used to achieve the required manipulator endpoint position and two

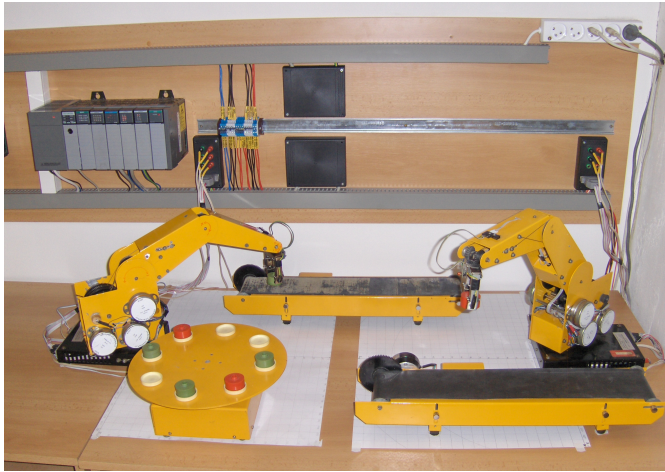


Fig.1. Experimental Workstation Model

are used for the grip orientation. Every arm is driven with six stepper motors and is equipped with two optical sensors and one pressure switch. Optical sensors are used to sense the transferred material in the grip and detect the initial position of the manipulator. The pressure switch signalizes that the material is grabbed with enough strength. The conveyors also contain optical sensors to sense the presence of the material.

The robotics Workstation control system [7] is based on the three-level distributed control system hierarchy (Figure 2). The technological level of control consists of single robotics workstation parts and a programmable logic controller SLC500 from Allen-Bradley, which controls the functionality of the whole workstation. The control program was built in RSLogix500 software. It is created in the way to control the movement of individual workstation parts and control manipulators through the kinematics model. The connection between the technological level and higher levels of DCS is realized thru the RS232 interface. The core of the supervisory level is presented by a local computer. The visualization software for local and remote supervisory control, the kinematics model software and RSQL, which is used to interconnect the technological level with information level, is installed on this PC. The local and remote visualization was realized through the use of SCADA/HMI software products from Rockwell Automation. The information level of control is based on Oracle 10g database platform, Oracle Application Server 10g and Oracle Business Intelligence 10g software components. The main part of the information level is the Oracle database, which saves technological data. The data can be than utilized to user requested form.

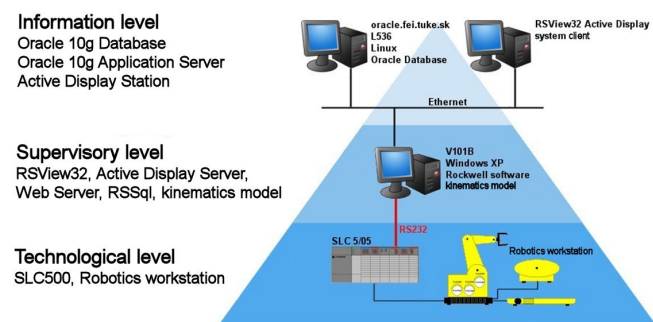


Fig.2. Robotics workstation control system hierarchy

#### IV. PRODUCTION LINE EXAMPLE

In this section will be shown a simple example of a production line composed from single parts of the robotics workstation, which task is to classify the material placed on the input palette 1. The specification of the technological process shown in Figure 3 is as follows.

- (1) The input palette 1 contains two types of pieces, material with white and black color arbitrarily arranged.
- (1) Manipulator 1 carries material from the palette 1 to the conveyor.
- (2) When the infrared sensor 1 detects the presence of the material the conveyor starts to move.
- (3) When the infrared sensor 2 detects the material the conveyor stops.
- (4) The camera senses the color of the material.
- (5) Manipulator 2 carries the material to palette 2 or palette 3 according to the scanned color.

The CPN model of the production line is shown in Figure 4. Table 1 provides a description of places and transitions involved. In the CPN it is possible to create one place, which will contain two kinds of tokens, representing the colors of the material placed on the input palette 1. This place is in the Petri net model marked as P1. The set of colours  $CB$  is defined as  $colset\ CB = with\ c / b$ . This means that places that are the type of  $CB$  can contain black or white tokens. The variable  $x$  is a type of  $CB$  ( $var\ x : CB$ ). The place P2 is a type of  $E$  ( $colset\ E = with\ e$ ), what indicates, that this place contains tokens without individuality. This marking is used because after firing T5 or T6 manipulator 1 can carry the next material on the conveyor.

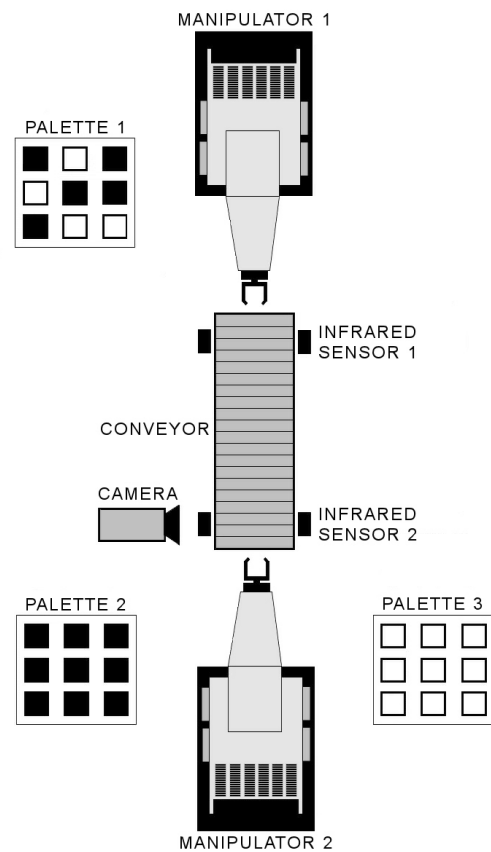


Fig.3. Production Line Example

V. CONCLUSION

This paper described a coloured Petri net model of a school manipulator workstation model on the technological level of control. The main task in the future is to create a model of the manipulator workstation considering the relations of individual function blocks in connection with the higher levels of the DCS in order to evaluate quality and diagnose the system.

REFERENCES

- [1] F. Kováč, "Distribúované riadiace systémy" Slovak University of Technology in Bratislava, 1998
- [2] H. J. Moon, "Performance Analysis and Design of a Communication Network for Industrial Automation", Ph.D. Thesis, Seoul National University, 1998
- [3] R. Zurawski, M. Zhou, "Petri nets in industrial applications: A Tutorial", IEEE Transactions on industrial electronics, vol. 41, No. 6, pp. 567 – 583, December 1994
- [4] K. Jensen, "A Brief Introduction to Coloured Petri Nets", Computer Science Department, University of Aarhus, Denmark
- [5] K. Jensen, "Coloured Petri Nets. Basic Concepts, Analysis Methods and Practical Use. Volume 1, Basic Concepts. Monographs in Theoretical Computer Science, Springer-Verlag, 2nd corrected printing 1997. ISBN: 3-540-60943-1.
- [6] K. Jensen, "Introduction to the Theoretical Aspects of Coloured Petri Nets", Article in Computer Science, Computer Science Department, Aarhus University, Denmark, August 28-30, 1999.
- [7] J. Chovaňák, "Riadenie a vizualizácia modelu sústavy manipulátorov na dispečerskej a informačnej úrovni riadenia", Diploma Thesis, Technical University of Košice, 2007

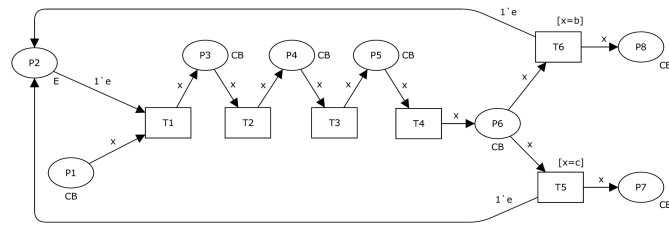


Fig.4. CPN model of the production line

TABLE I  
INTERPRETATION OF PLACES AND TRANSITIONS

Place	Interpretation
P1	Material is ready on the palette 1
P2	Manipulator 1 is ready in the initial position
P3	Infrared sensor 1 senses the presence of the material
P4	Infrared sensor 2 senses the presence of the material
P5	Camera ready
P6	Manipulator 2 ready in the initial position
P7	The black material carried to the palette 2
P8	The white material carried to the palette 3
Transition	Interpretation
T1	Manipulator 1 carries the material to the conveyor
T2	Conveyor start
T3	Conveyor stop
T4	Camera color detection
T5	Manipulator 2 carries the material to the palette 2
T6	Manipulator 2 carries the material to the palette 3

The coloured Petri net model was created in the CPN Tools software, which is a great tool for editing, simulating and analysing coloured Petri nets. The production line CPN model simulation is shown in Figure 5 (initial marking) and Figure 6 (end of the simulation). As we can see from the initial marking, where place P1 holds five pieces of black material and four pieces of white material, after the simulation the material was classified and placed on the correct palette.

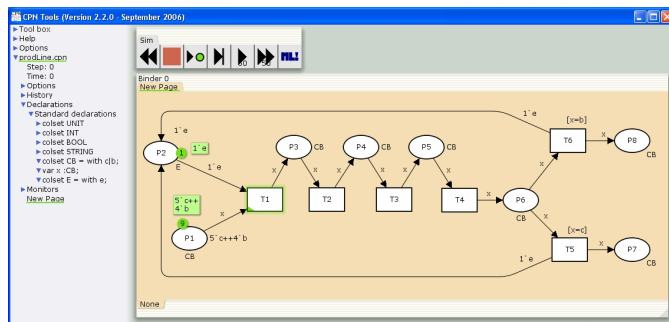


Fig.5. Production line CPN model initial state

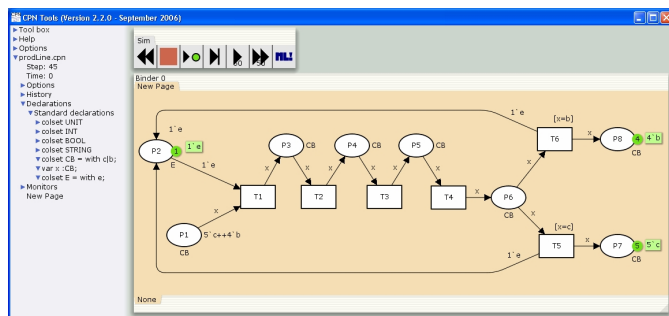


Fig.6. Production line CPN model after simulation

# Focused magnet for drug targeting

<sup>1</sup>L. Jancurová

<sup>1</sup>Dept. Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

<sup>1</sup>lucia.jancurova@cern.ch

**Abstract**—The investigation of biocompatible magnetic nano-carrier systems, e.g. magnetic liquids such as ferrofluids, are included in current research on methods to target chemotherapy drugs in the human body. The aim of our work was to construct a focused magnet, which enables to achieve maximal magnetic force in deeper position, to map its magnetic field and to find the adhesion condition for a magnetic fluid drop in magnetic field with obtained design.

**Keywords**—magnetitic drug targeting, magnetic fluid, focused magnet

The difference between success and failure of chemotherapy depends not only on the drug itself but also on how it is delivered to its target. One of the major problems in pharmacotherapy is the delivery of drugs to a specific location and maintenance of its location for the desired length of time. Because of the relatively non-specific action of chemotherapeutic agents, there is almost always some toxicity to normal tissues. Therefore, it is of great importance to be able to selectively target the magnetically labelled drug to the tumor target as precisely as possible, to reduce resulting systemic toxic side effects from generalized systemic distribution and to be able to use a much smaller dose, which would further lead to a reduction of toxicity. The method of magnetic drug targeting is dependent on physical properties, concentration and amount of applied nanoparticles, on type of binding of the drugs, on the physiological parameters of the patient and of course on magnetic force, which is defined by its field and field gradient [1].

Guided transport of biologically active substances to the target organ allows creating an optimum therapeutic concentration of the drug in the desired part of organism, while keeping the total injected dose low [2-4]. Current research on methods to target chemotherapy drugs in the human body includes the investigation of biocompatible magnetic nano-carrier systems, e.g., magnetic liquids such as ferrofluids. The use of biocompatible magnetic fluid as potential drug carrier appears to be a promising technique. Due to their superparamagnetic properties the magnetic fluid drops can be precisely transported, positioned and controlled in desirable parts of blood vessels or hollow organs with the help of an external magnetic field. The motion of magnetic drop within the body is controlled by the combination of magnetic force and a hemodynamic drag force due to blood flow. The models which investigate the interaction of an external magnetic field with blood flow containing a magnetic carrier substance are based on the Maxwell and Navier-Stokes equations, where a static magnetic field is coupled to fluid flow. This is achieved by adding a magnetic volume force to the Navier Stokes equations, which stems from the solution of magnetic field problem [5]. In order to effectively overcome

the influence of blood flow the magnetic force must be larger than the drag force. The conditions for holding a magnetic fluid drop on a blood vessel wall were investigated by Voltairas et al.[6]. In this work the non-uniformity of considered magnetic field as higher only close to the magnetic pole, what was regarded as a major technical problem that has to be resolved in order for the drug targeting to remain essentially non-invasive. The aim of our work was to construct a focused magnet, which enables to achieve maximal magnetic force in deeper position, to map its magnetic field and to find the adhesion condition for a magnetic fluid drop in magnetic field with obtained design.

Voltairas et al. [6] presented a self-consistent ferrohydrodynamic theory of magnetic drug targeting and examined a model case to account for adhesion.

They obtained an upper bound of the mean blood flow velocity as a function of the applied magnetic field, which was considered to be produced by a point source located outside the body at  $x = -\delta$ ,  $y = 0$ ,  $z = \zeta$ , ( $\delta, \zeta > 0$ , non-uniformity higher only close to magnetic pole) and had the form

$$\vec{H} = \frac{m(\vec{r} + \delta \hat{e}_x - \zeta \hat{e}_z)}{(r^2 + \delta^2 + \zeta^2 + \sqrt{\delta^2 x^2 - \zeta^2 z^2})^{3/2}} \hat{i} \quad (1)$$

where  $\mathbf{m}$  is the magnetic dipole moment. The magnetic point source was oriented at an angle

$$\omega = \arcsin\left(\frac{\zeta}{\delta}\right) \quad (2)$$

with respect to the x-axis.

The found adhesion condition in dimensionless form read

$$\frac{1}{B_m} = \frac{\chi}{4S_0} \int_{S_1} \int (h^2 + h_n^2) dS \quad (3)$$

where

$$B_m = \frac{\mu_0 H_0^2 R}{\gamma} \quad (4)$$

is the magnetic bond number with  $R$  being the radius of the magnetic drop and

$$h = \frac{H}{H_0}, \quad h_n = \frac{H_n}{H_0}$$

$$H_0 = \frac{m}{\delta^2},$$

$$S_0 = 2\pi R^2 \quad (5)$$

An additional global condition, taking into account the deformation of the magnetic drop due to the blood flow, was derived in form

$$V_m = \frac{\chi}{2\beta S_0} \int_{S_1} \int \left[ (h^2 - (1+2\chi) h_n^2) \hat{n}_z + 2\chi(1+\chi) h_n h_z \right] dS \quad (6)$$

where

$$V_m = \frac{\eta_2 u_0}{\mu_0 H_0^2 R} \quad (7)$$

is the dimensionless velocity and

$$\beta = \frac{\gamma_v - 1}{\gamma_v + 1}$$

$$\gamma_v = \frac{\eta_2}{\eta_1}$$

$$h_z = \frac{H_z}{H_0} \quad (8)$$

Here  $\eta_2$  is the blood viscosity,  $\eta_1$  is the viscosity of magnetic fluid and  $u_0$  is the mean blood flow velocity. Thus, instead of one adhesion condition, Voltairs et. al [6] obtained two equations (3) and (6) and the dependence of blood flow velocity on the applied magnetic field was parameterized as

$$B_m = B_m(R, \delta, \chi, \omega) \quad (9)$$

and

$$V_m = V_m(R, \delta, \chi, \omega, \gamma_0) \quad (10)$$

The obtained law  $V_m = V_m(B_m)$  gives an upper bound of the mean blood flow velocity, at which the applied magnetic field is able to capture a magnetic drug drop on the blood vessel wall.

To achieve a proper non-uniform magnetic field, able to localize a magnetic drop inside the body, two types of magnets could be used - permanent magnets or electromagnets. The permanent magnets generate magnetic fields, which are rather weak for the trapping of magnetic drop in a bulk vessel. The

electromagnets generate stronger magnetic fields, however, they need an outer source of electrical current and they produce the Joule heat. Our aim was to develop an arrangement of permanent magnets which could generate higher induction and gradient of magnetic field than classical magnets with simple geometry.

This theoretical background was consecutively used for suggestion design and construction of special focused magnet. It is evident, that manufacturing such magnet from one piece would be very problematic. Nevertheless, some approach which realizes the principle at least in reasonable approximation should be viable. We could apply the information that the direction of magnetization is parallel to the straight lines passing the focus. It means that the magnet could be composed of pyramids with peaks in focus  $F$ ; the direction of the magnetization in the pyramids should be parallel. Following the above mentioned considerations and taking into account the technological simplicity, we proposed the construction of compound focused magnet. Its cross section is schematically drawn in Fig. 1

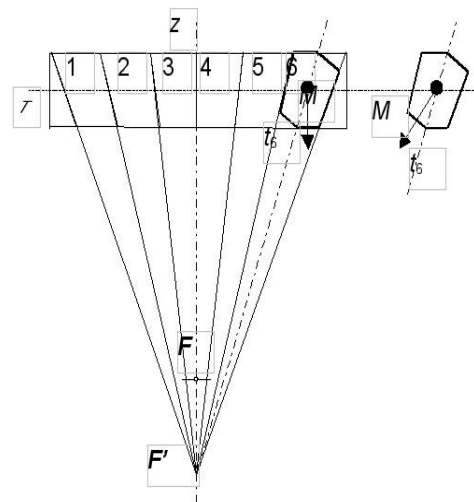


Figure 1: The crossed section of the focused magnet

The original FeNdB magnet had a form of rectangular prism (40 x 40 x 10 mm), the preferred direction of magnetization was perpendicular to its greatest side. Before the magnetization the magnet was cut into six prism by an electro-spark cutter, according to the scheme shown in Fig. 1. Moreover, each prism was shaped on such a way, that its cross section became symmetric according to the axis crossing its center of mass and peak  $F'$ . This modification is shown for the prism No.6. Then the prisms were turned over 180° around the axes  $t_i$ ;  $i=1-6$  and glued to their neighbours. After the gluing procedure to six prisms followed, according to the same scheme. The only difference was, that the glued magnet was turned by right angle around axis  $z$  before the new cutting, thus a checked structure of the magnet was obtained. The turning and gluing of the prisms was repeated and finally the compound intermediate magnet was obtained, with magnetization of each part directed into approximately one point on the axis  $z$ . This point lies in the middle of the distance between the plane  $\tau$  and peak  $F'$  (in plane  $\tau$  the centers of mass of the parts of compound magnet lie). We remark that point  $F$  lies approximately in  $3/4$  of the distance  $\tau-F'$ . After the second gluing the intermediate magnets were enclosed into a brass mantle and magnetized in a homogeneous magnetic field 15 T. This way the focused magnet, designed for the magnetic

targeted drug delivery, was obtained.

Consecutively we have tested the constructed magnet from point of magnetic field induction and magnetic field gradient, respectively. The magnetic field of manufactured focused magnet was measured by 3D Hall probe. The magnetic field profile in axis z is shown on Fig. 2.

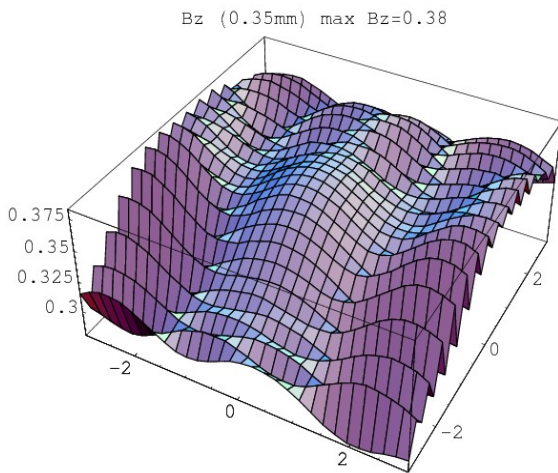


Figure 2: The profile of magnetic field generated by focused magnet

The value of magnetic field at focus was estimated to be 0.38 T and magnetic field gradient was estimated to be 110 T/m respectively. In effort to test the ability of the constructed magnet the found profile of its magnetic field was used in the numerical calculation following the Voltairas et al. model [6] i.e. to find upper bound velocity to hold magnetic fluid drop in carotid artery - Table 1

$B_0$ [T]	0.38
$u_0$ [m/s] (computed)	0.84
$u_{exp}$ [m/s] (experiment)	0.1- 0.6
$F_m$ [kN/m <sup>3</sup> ]	105.03
$dB/dz$ [T/m]	110T/m
$M$ [mT]	1

Table 1: The model – comparison with experiment for carotid artery.

where comparison with experimental data are given too. As it is seen used magnetic field geometry can hold magnetic fluid drop (saturation magnetization of MF was 1mT i.e. 10 Gauss) in carotid artery. So it can be said, that using our specially focused magnet a higher magnetic field as well as its gradient can be achieved in deeper position, what could enable to the non-invasivity of the magnetic drug targeting procedure.

In summary, a focused magnet consisting of 36 prisms with pyramidal shape was manufactured, generating higher magnetic field and higher magnetic field gradient as compared with classical prism. The magnetic field of the focused magnet was mapped and its profile was used in numerical calculations, which yielded the upper bound of the mean blood flow velocity, at which the applied magnetic field is able to capture a magnetic drug drop on the blood vessel wall. The obtained

results verified the ability of the magnet to generate a sufficient magnetic force in deeper position, what could contribute to the non-invasivity of the magnetic drug targeting procedure.

ACKNOWLEDGMENT

This work was supported by JINR Dubna protocol No: 3920-6-09/10

REFERENCES

[1] T. Neuberger, B. Schöpf, H. Hofmann, M. Hofmann, B. Rechenberg, *J.Magn.Magn.Mater.*293 (2005) 483.  
 [2] A. A. Kuznetsov, V.I. Filipov, O.A. Kuznetsov, V.G. Gerlivanov, E.K. Dobrinsky and S.I Malashin, *J.Magn.Magn.Mater.*194 (1999) 22.  
 [3] E.K. Ruuge, A.N. Rusetski, *J.Magn.Magn.Mater.*122 (1993) 335.  
 [4] U.O. Häfeli, J.G. Pauer, *J.Magn.Magn.Mater.*194 (1999) 76.  
 [5] R. Ganguly, A.P. Gaiind, S. Sen, I.K. Puri, *J.Magn.Magn.Mater.*289 (2005) 331.  
 [6] P.A. Voltairas, D.I. Fotiadis and L.K. Michalis, *J.Biomech.* 35 (2002) 813.. 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.

# Data Analysis on Grid using AliEn

<sup>1</sup>Lucia JANCUROVÁ, <sup>2</sup>Martin VAĽA

<sup>1</sup>Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

<sup>2</sup>Institute of Experimental Physics, Slovak Academy of Sciences, Košice, Slovak Republic

<sup>1</sup>lucia.jancurova@cern.ch, <sup>2</sup>martin.vala@cern.ch

**Abstract**—Due to start-up in September 2009, the Large Hadron Collider (LHC) will provide collisions at the highest energies ever observed in laboratory conditions and physicists are eager to see what they will reveal. Four huge detectors – ALICE, ATLAS, CMS, LHCb – will observe the collisions so that the physicists can explore new territory in matter, energy, space, and time.

These experiments will produce huge amount of data. To save and analyse this amount of data, it is needed to have a lot of storage space (disks, tapes, ...) and many computers (farms). By connecting many computing farms together, it was created array of farms, so called Grid.

Each experiment has its own modified version of grid software (gLite). For example, ALICE experiment uses grid software called AliEn and for grid monitoring uses Monalisa software.

We created class `TAlienJDL` which can generate `jdl` file using `root` macro and then run AliEn job directly from `Root`.

We also created class `TAlienPackage`. It is the class which allows us to copy program (`aliproot`, `root`, ...) from AliEn to our local desktop. So we can test our own macros on local pc before sending them to the Grid.

**Keywords**—ALICE, AliEn, Grid, LHC.

## I. INTRODUCTION

The Large Hadron Collider (LHC) is being built in a circular tunnel 27 km in circumference. The tunnel is buried around 50 to 175 m. underground. Located between the Jura mountain range in France and Lake Geneva in Switzerland, the tunnel was built in the 1980s for the previous big accelerator, the Large Electron-Positron collider (LEP).

The LHC will produce head-on collision between two beams of particles of the same kind, either protons or lead ions. The beams will be created in CERN's [1] existing chain of accelerators and then injected into the LHC, where they will travel through a vacuum comparable to outer space. Superconducting magnets operating at extremely low temperatures will guide the beams around the ring.

The first beams were circulated successfully on 10th September 2008. Due to switch on in 2009, the LHC will provide collisions at the highest energies ever observed in laboratory conditions and physicists are eager to see what they will reveal. Four huge detectors – ALICE, ATLAS, CMS, LHCb – will observe the collisions so that the physicists can explore new territory in matter, energy, space, and time.

## II. THE ROOT SUPPORT OF THE ALIEN

The **ALICE** (A Large Ion Collider Experiment) Collaboration is building a dedicated heavy-ion detector to exploit the unique physics potential of nucleus-nucleus interactions at LHC energies.

The **ATLAS** (A Toroidal LHC Apparatus) is a particle physics experiment at the LHC at CERN. Starting later in 2008, the ATLAS detector will search for new discoveries in the head-on collisions of protons of extraordinarily high energy.

The **CMS** (Compact Muon Solenoid) experiment uses a general-purpose detector to investigate a wide range of physics, including the search for the Higgs boson, extra dimensions, and particles that could make up dark matter.

The **LHCb** (Large Hadron Collider beauty) experiment will help us to understand why we live in a Universe that appears to be composed almost entirely of matter, but no antimatter.

### A. The Grid

The software infrastructure providing automatic distribution of analysis tasks to distributed resources is called GRID middleware. Grid computing in general is a special type of parallel computing which relies on complete computer connected to a network by a conventional network interface. This is in contrast to the traditional notion on a supercomputer, which has many processors connected by a local high-speed computer bus. This will link tens of thousands of computers worldwide to create a vast global computing resource for the LHC experiments.

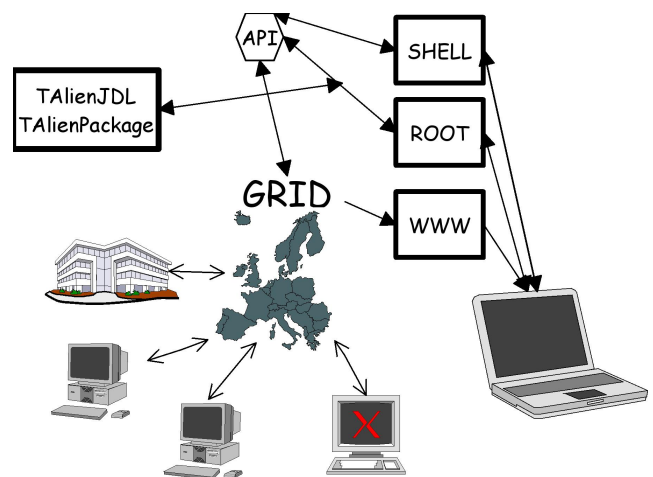


Fig. 1. Grid dependences

### MonAlisa for the ALICE experiment

Computing Center's providing resource for the ALICE experiment. It lets numerous computing resources appear like a single system, executes analysis tasks on suitable resources and retrieves the results.

## AliEn – ALICE Environment

AliEn [2], [3] is the Grid middleware developed by the ALICE experiment. It provides data and job management, splitting of time-consuming analyses. It is accessible via ROOT, thus the same analysis macros can be executed locally (e.g. to verify the results) and on the Grid with a much larger number of input files. More than 3650 simulation and reconstruction jobs were run concurrently at 15 computing centers (sites) in tests for the upcoming physics data challenge.

The next code corresponds with the programs shown in the Figure 1: Class which creates JDL files for the AliEn middleware

```
//_____
Bool_t TAlienJDL::SubmitTest (Bool_t doSubmit)
{
    // Tests the submission of a simple job.

    Info ("SubmitTest", "submitting test job
        /bin/date");

    if ( !gGrid )
    {
        Error ("SubmitTest",
            "you must have a proper GRID
            environment initialized");
        return kFALSE;
    }

    Clear();
    SetExecutable ("batch.sh");
    SetSplitMode ("se",5,100);
    SetJobTag ("comment:AliEn tutorial");
    AddToRequirements ("member(other
        GridPartitions,\"Analysis\")");
    AddToPackages ("APISCONFIG","V2.2");
    AddToPackages ("ROOT","v5.15.04",
        "VO_ALICE");

    SetTTL (30000);
    SetInputDataList ("wn.xml");
    SetInputDataListFormat
        ("merge:/alice/cern.ch/user/l
        /ljancuro/test/ESD/global.xml");
    AddToInputDataCollection
        ("LF:/alice/cern.ch/user/l
        /ljancuro/test/ESD/global.xml,nodownload");

    AddToInputSandbox ("LF:/alice/cern.ch/user
        /l/ljancuro/test/ESD/runProcess.C");
    AddToInputSandbox ("LF:/alice/cern.ch/user
        /l/ljancuro/test/ESD/demoBatch.C");
    AddToInputSandbox ("LF:/alice/cern.ch/user
        /l/ljancuro/test/ESD/ESD.par");
    AddToInputSandbox ("LF:/alice/cern.ch/user
        /l/ljancuro/test/ESD/ANALYSIS.par");
    AddToInputSandbox ("LF:/alice/cern.ch/user
        /l/ljancuro/test/ESD
        /AliAnalysisTaskPt.h");
    AddToInputSandbox ("LF:/alice/cern.ch/user
        /l/ljancuro/test/ESD
        /AliAnalysisTaskPt.cxx");
    AddToInputSandbox ("LF:/alice/cern.ch/user
        /l/ljancuro/test/ESD/demoBatch.C");

    AddToOutputArchive ("log_archive:stdout,
        stderr@Alice::CERN::se");
    AddToOutputArchive ("root_archive.zip:*.*
        root@Alice::CERN::se");

    SetEMail ("Lucia.Jancurova@cern.ch");
}
```

```
AddToOutputSandbox ("different.
        root@ALICE::CERN::SE");
SetValidationCommand ("DONE");
SetSplitArguments ("arg1 arg2");

AddToMerge();
AddToMerge ("hist.root",
    "/alice/jdl/mergerootfiles.jdl",
    "histAll.root" );

if ( doSubmit == kTRUE )
{
    TGridJob* job = gGrid->Submit ( Generate() );
    printf ("Sending:\n%s\n",Generate().Data() );
    if ( job == 0 )
    {
        Error ("SubmitTest", "submitting failed");
        return kFALSE;
    }
    else
    {
        printf ("JDL is :\n%s\n",Generate().Data());
    }
    return kTRUE;
}
}
```

The next code corresponds with the programs shown in the Figure 1:

New class providing package management functionality like the AliEn Package Management System. Allows to setup software packages on a local desktop like in the GRID environment and to execute the contained programs. Registered Dependencies are automatically resolved and missing packages are automatically installed. Currently there is no support for 'source' packages. The desired platform has to be specified in the constructor. The default constructor takes packages from the global package section in AliEn. If you want to install a user package, you have to set the AliEn package directory to your local package directory using:

```
package->SetAliEnMainPackageDir("/alice/cern.ch/
user/.../packages")
```

```
ClassImp ( TAlienPackage );

TAlienPackage::TAlienPackage()
    :fInstallList ( 0 ),
    fDebugLevel ( 0 )
{
    if ( GetDebugLevel() >2 )
        cout <<"Default constructor"<< endl;
    fName= "ROOT";
    fVersion="v5-16-00";
    fPlatform="Linux-i686";
    fAliEnMainPackageDir="/alice/packages";
    fInstallationDirectory="/tmp";
    fPostInstallCommand="post_install";

    if ( !gGrid )
        gGrid = TGrid::Connect ( "alien://" );

    if ( !fInstallList )
        fInstallList = new TList();
}
...

Bool_t TAlienPackage::Install()
{
}
```



```
fInstallList->Clear();
if ( GetDebugLevel() >1 )
    cout <<"installing"<< endl;

if (CheckDirectories
    (fName,fVersion) == kFALSE ) return kFALSE;

if (CheckDependencies() == kFALSE )
    return kFALSE;

if (InstallAllPackages() == kFALSE )
    return kFALSE;

return kTRUE;
}
```

### III. CONCLUSION

The packages described above enable user to run programs on his local computer. The program `TAlienPackages` will determine the software installed on the local computer with dependences, and will install all necessary program packages.

### ACKNOWLEDGMENT

This work was supported by the Slovak Academy of Sciences (grant No. 6166) and Science and Technology Assistance Agency within the contract No. APVV-99-0265.

### REFERENCES

- [1] "www.cern.ch"
- [2] "www.alien.cern.ch"
- [3] A. J. Peters and Feichtinger, "Authorization of data access in distributed storage systems," internal Report, Oct. 2003.

# Mobility management in VoIP networks: Intelligent networks vs. Intelligent applications

*Jozef JANITOR, Peter FECILAK*

Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

{firstname.lastname}@tuke.sk

**Abstract** — Wireless Internet access has recently become a very general form of network access. Technologies like WIFI (802.11\*), 3G UMTS and 4G networks can provide enough bandwidth and reliability to implement realtime IP services from “wire” data networks to wireless environments. Web browsing, emailing, or watching online videos is not bonded to a physical computer anymore, but is accessible through different kind of mobile devices (PDAs, Mobile phones, etc.). Although wireless technologies are ready to use, and work well in areas covered with the signal of particular wireless technology, there are still missing standardized methods of doing transparent handovers between different wireless and wire access technologies. Web browsing, instant messaging and emails are network services that are generally not sensitive to a network outage caused by roaming from one network access technology to another. On the other hand, realtime network services like Voice over IP and/or Video over IP are highly sensitive even to a short network access outage that might easily lead to service interruption. The paper presents some technologies that can be used to provide uninterrupted services while roaming between different network access technologies.

**Keywords**— Mobile IP, SIP, VPN, NGN, UMA, IMS, 3G, WIFI.

## I. INTRODUCTION

Internet Protocol (IP) is used for communicating data between endpoints in IP based, packet switches data networks. IP protocol uses IP addresses to logically identify endpoints in their current location. Usually each physical location and network access technology defines its own range of IP addresses, which are used by routers to find the best path between communicating parties. When the endpoint’s IP address is changed, all active communication channels that were previously active, must be reactivated, thus causing a service interruption. Services like web browsing, instant messaging, etc. can easily recover from this interruption just by reconnecting to a server. The end user might sometimes need to refresh the page displayed in a web browser, or restart an interrupted file download, but the whole user experience might be still at high quality. On the other hand, when services like Voice over IP and Video over IP are interrupted even just for a short period of time, the user experience is getting worst.

When roaming between different networks or networking technologies, the IP address is usually changed. That means, for example, when a device is connected to a 3G wireless network, and creates a voice or video over IP (V/VoIP) call

through that connection, but later finds a new and better network access technology available (WIFI, Ethernet wire, etc.), it cannot just roam because it would cause an IP address change which leads to a call termination. The device cannot even roam to an optional network access technology, when the signal from the 3G network is dropping in the current area, without terminating a V/VoIP call.

To overcome this issue, two options exist:

- Intelligent Networks - it is necessary to maintain the assigned IP address of a device to survive a handover roaming between different network access technologies.
- Intelligent Applications - use application layer protocols to detect network changes and reroute data flows to new addresses.

## II. INTELLIGENT NETWORKS

Intelligent Networks can deal with the change of network access technology, simply by implementing mobility support directly into the network. Networking devices and protocols can detect roaming actions of end devices, and reroute the traffic flows between them to new paths. When using WIFI, transparent roaming between wireless access points can be easily achieved by putting all the access points to the same VLAN. There is a short service interruption caused by disassociating from the old access point and associating to the new one, but as the end device remains in the same VLAN, there is no need to change its IP address. On the other hand, when roaming between different networks, the IP address must be renewed.

Protocols like Mobile IP [1] were developed to provide transparent mobility for end users and end devices between different networks. When an end-device is turned on, it registers itself to a Home Server and creates a VPN\_like tunnel that is used to forward data packets. The tunnel has its own Mobile IP address and primarily it is used to forward packets to the end device from other hosts. Whenever it is possible, the end device communicates with other hosts directly while using the assigned Mobile IP as a source address, thus skipping the tunnel. That decreases the total delay caused by forwarding packets through a tunnel. On the other hand, this method cannot be used when the transporting network checks if the source IP address corresponds to IP addresses that are allocated to that area. As the end device uses its Mobile IP address as a source address to communicate

with other hosts, these packets are usually dropped as spoofed packets by security filters on routers.

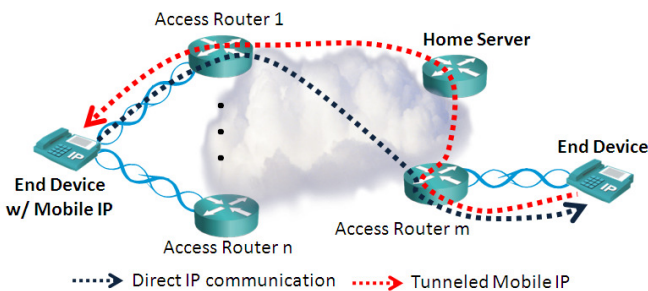


Fig. 1. Mobile IP packet flow with a single tunnel path

Reverse tunnels offer a solution by tunneling not only ingress end device traffic, but egress as well. This solution adds additional overhead to the traffic, as now both ways of a communication channel are forwarded through a tunnel, that creates from a Mobile IP a true VPN tunneled technology. On the other hand, with reverse tunnels, Mobile IP works in almost every network. One of the disadvantages is a need to have a Home Server that terminates VPN tunnels, redistributes Mobile IP addresses into standard IP routing protocols, etc. Scalability is therefore limited by resources on a Home Server. Also, end devices must support Mobile IP protocol and have a Mobile IP client installed. Other issues are on the security side of this solution, as in reverse tunneled mode the Home Server acts as a Man In the Middle person in the communication channel.

### III. INTELLIGENT APPLICATIONS

Applications can detect network changes and programmatically adjust settings to new network conditions. Operating systems can inform applications about roaming to a new network access technology by an API call that can start the adjustment process. Realtime communication applications usually divide the signaling channel from the actual data, or payload channel. The payload (voice/video) is transported over the IP network encapsulated in RTP and UDP packets. As in many types of datagram based multimedia communication, before the communication channel for payload is established, signaling protocols are used to create an agreement to define on which IP addresses and ports the payload communication is going to happen. Session Initiation Protocol (SIP) [3] and Session Description Protocol (SDP) [4] are one of the most common protocols used as signaling protocols. SIP signaling is used to create, maintain, manage, and terminate calls. The SIP architecture defines the following components:

- User Agent Client (UAC), User Agent Server (UAS)
- Proxy
- Redirect Server
- Location Server

UAC and UAS are implemented as SIP terminals – end devices – IP Phones. The SIP Proxy server is used to forward the signaling between two or more SIP terminals. It usually implements itself the functionality of the SIP Redirect and the SIP Location server too. The SIP Redirect server receives and redirects client requests to real address of a called SIP terminal. The SIP Redirect server uses information stored in

Location server’s database to find the current logical location address (IP address, port, etc.) of a called SIP terminal. To keep the location database up to date with fresh location information, SIP Terminals periodically update information about their current logical location address by sending SIP REGISTER messages to the location server.

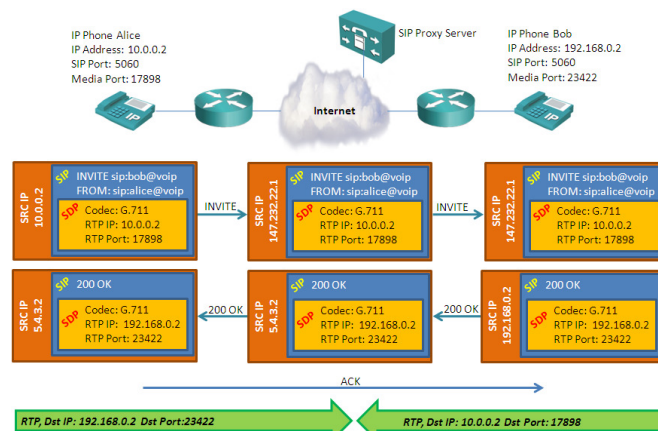


Fig. 2. VVoIP call establishment with SIP signaling RTP payload

SIP itself does not provide information about the payload transport. Rather it uses SDP to pass this information between end devices. SDP protocol is encapsulated in SIP signaling messages.

To setup a call, a SIP INVITE message (Table 1) is sent towards the called SIP Terminal, containing information about a called party and an SDP part. The SDP part provides information about various types of multimedia sessions supported by the calling party and its addresses (IP address, port number, etc.) that are going to be used to create a payload data flows. Each SIP call is uniquely identified with its unique Call-ID. After a call is successfully established, the Call-ID information is used to identify and manage the active call. When a user terminates a call by pressing the HangUp button, a SIP BYE message with the identifying Call-ID field is sent to the called party.

```

INVITE sip:7400@cml.tuke.sk SIP/2.0
Via: SIP/2.0/UDP 192.168.192.2:5070;
To: <sip:7400@cml.tuke.sk>
From: "Jozef Janitor" <sip:jozjan@cml.tuke.sk>
Call-ID: JJ-90ED9032C2E1@192.168.192.2
CSeq: 532 INVITE
Max-Forwards: 20
User-Agent: Express Talk 2.02
Contact: <sip:jozjan@192.168.192.2:5070>
Allow: INVITE, ACK, CANCEL, OPTIONS, BYE
Supported: replaces
Content-Type: application/sdp
Content-Length: 312

v=0
o=- 827905653 827905665 IN IP4 192.168.192.2
c=IN IP4 192.168.192.2
t=0 0
m=audio 8000 RTP/AVP 0 8 96 3 13 101
a=rtpmap:0 PCMU/8000
    
```

Fig. 3. SIP INVITE encapsulating SDP protocol

In similar way as the user can terminate the call, he can transfer it to a different phone number by sending new SIP INVITE message with the Call-ID identifying the current call. When a new device contacts the called party with the same Call-ID, he can get access to the active session, and thus gain control over the call. This feature can be used to overcome a handover issue caused by roaming between different network access technologies.

When a client detects a change on its IP address and therefore payload's packet loss, it can reactivate the payload flow simply by sending a new SIP INVITE message, containing the original Call-ID field to the called party. That looks to the called party as a call transfer, so it will continue sending the payload data packets to a new address that was learned from the SIP INVITE encapsulated SDP information.

Even though the call will be interrupted for a short time, because of active packet loss of a payload that is being sent to the previous IP address and port number, it will get restored immediately as the called party will update the payload flow's new destination address learned by a new SIP INVITE message.

To being able to receive new calls, the client must also update its location information at the SIP Location server by sending a new SIP REGISTER message.

#### IV. CONCLUSION

The paper presents two different approaches to handle the issue of a network access technology handover.

The first method uses techniques provided by an Intelligent Network, that can reroute, or tunnel packets between end devices as it is in a VPN tunnel. This method requires additional resources on the Service Provider's network, like a Home Server router that terminates and manages the Mobile IP tunnels. Every end device node requires to have a client application that initializes the Mobile IP session and maintains it through a call. One of the highest disadvantage when considering realtime multimedia sessions, is the additional payload and delay that is caused by the tunneling methods of this solution.

The second method uses techniques found in Intelligent Applications, that can detect network change and dynamically adapt to new conditions. The paper presents the usage of a SIP call transfer method to overcome of the issue caused by the handover. This method does not require special changes in the Service Provider's network, it generates no additional overhead and delay to the payload and does not need installation of additional applications to the end device. However, the handover management logic must be implemented in the calling application. The implementation on the other hand must not be difficult as it can reuse standard methods like call transfer.

Both of the presented methods can be used to overcome the handover caused issues. The second one – Intelligent Applications – should be preferred over the Intelligent Network methods as they add no additional complexity to the architecture, but require some changes in the application. Whenever the application that is running on the end device is unable to refresh its communication channels (e.g. uses TCP as a transport protocol, etc.) without terminating the current

session, the first option – Intelligent Networks – can be a good option to keep the active session up and running while roaming between different network access technologies.

#### APPENDIX

The term “Network access technology” is used in the paper as an identifier to different technologies like WIFI, 3G, 4G, Cable Ethernet, etc.

#### ACKNOWLEDGMENT

We would like express our greatest thank to the Computer Networks Laboratory at DCI FEEI TU Kosice Slovakia for the support of this work.

The paper was partially prepared within the project “Methods of multimedia information effective transmission”, No. 1/0525/08 with the support of VEGA agency and the project EDINET - E-learning in Distributed Data Network Laboratory, 134608-LLP-2007-1-FI-ERASMUS-EVC.

#### REFERENCES

- [1] RFC2002 - IP Mobility Support
- [2] S. Raab, *Mobile IP Technology and Applications*. Cisco Press, 2005.
- [3] RFC3261 - SIP: Session Initiation Protocol
- [4] RFC 2327 - SDP: Session Description Protocol

# Image acquisition of cell nuclei in micro-axial tomography

<sup>1</sup>Vladimír Jeleň

<sup>1</sup>Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

<sup>1</sup>vladimir.jelen@tuke.sk

**Abstract**—This article deals with process of image acquisition of data about cells and they nuclei with the aid of microaxial tomography. It interprets hardware and software necessary to collect the same type information and some ordinary aberrations in optical microscopy explains.

**Keywords**—Assignment, automated microscopy, image fusion, image registration, micro-axial tomography, resolution improvement.

## I. INTRODUCTION

The task was to create an application, which would enable observation of cells from all sides and create 3D model of them for further processing. It requires multidisciplinary access from areas of biology, informatics and mechatronics.

The idea of observing an object from several viewing angles and further reconstruction of this object in 3D region is not new. The first attempts are recorded already in 1975 and described in [3]. The microscope slide was tilted below microscope, but it enabled tilt slide maximum about 30°. This disadvantage can be removed by placing cells to capillary, where they can be observed in 360° area. It brought several problems. Capillary brought two rounded walls into the way of light ray and that causes its breaking, what practically stands that spotted object is deformed. Problem was solved by cell attached to glass fibre [5].

Similar device was created at Masaryk University in Brno described in [2]. This article contains procedure which inspired our work. Glass capillary was better to use for our work, despite the fact that capillary have more disadvantages.

It was needed to create a hardware, which should be possible to fix in desk microscope and at the same time had to be stable for manipulation. It had to allow rotation of samples and at the same time simple manipulation with them. This device is called adapter on gripping glass capillary, or accordingly micro-axial tomograph and in the following parts of this article it will be closely explained.

Further, it was needed to solve manner arrangements and saving biological samples, so that samples verify required optical facilities. The problem was in the optical aberration caused current traverse of optic ray across regions with

different optic density, what caused his deformation and then picture impairment.

## II. TILTING HARDWARE

### A. Tilting tools requirements

Registration method described here expects that the biological material will be placed in the capillary, which will be in the process of image acquisition, preferably in the steadiest attachment, so that oscillations in the axis x, y and z they will be minimum. Therefore it was necessary to create a device, which snaps capillary and lets it rotate without having big oscillations and at the same time lets it place on microscope stage and consequential analysis. This could be a problem, because between desk microscope and object lens is workspace with low height. Of course, it depends on the type of used microscope and techniques of microscopy.

For the measure was used bifocal microscope and fluorescence microscope.

Other problem was how to solve the rotation of the capillary. For this purpose the stepper motor was regulated over USB port.

Image scanning was provided with standard webcam attached accordingly through the USB port.

### B. Microscope

Bifocal microscope for our experiment was changed only minimally. For finding the best fix of webcam, it was took out one right ocular and replaced by webcam. However, it gave bad results. Better results were seen, when webcam was placed on an ocular, it was a simulation for humane perception. But now we got a problem here, how to fix it. Solution brought little plastic part clutch in form ring into which from inside is positioned camera and it was attached from the external side toward rubber ring of ocular.

### C. Camera

More types of optical cameras and webcams were used in experiments for image acquisition. It was applied for fluorescence microscope in the combination with camera most

frequently. Standard webcam was used for experiments on bifocal microscope, fixed through the medium simple plastic clutch, right on the one of ocular's microscope. In this way the picture was transferred right to the computer, where it was subsequently analyzed by software.

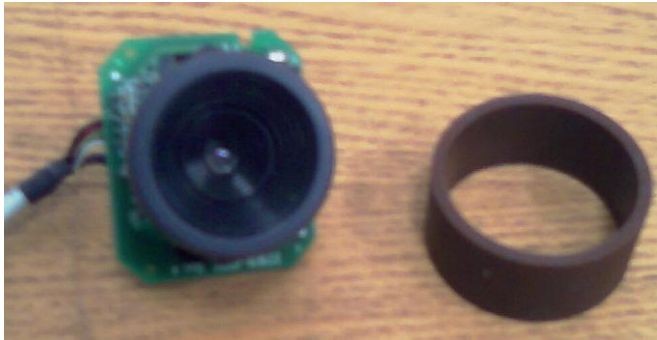


Fig. 1. Camera and adapter used to connect camera to microscope. It was necessary to create a ring adapter because of stabilizing camera to ocular of microscope. This solution enabled potential zoom of camera.

#### D. Mounting adapter

Adapter to grip glass capillary was made of aluminium and enables to grip one capillary and at the same time grip stepper board which rotates through the medium elastic clutch.

This device has been mounted from aluminium body (which can be positioned on working object stage), two steel letters of saw (scoring which help stabilize capillary in place), cover stability strip of tin (stabilizes capillary in place), two magnets (allowing the vertical adjustment positions of saw letters and at the same time cover stability strip of tin) and adjustable holder on step motor.

While construction the adapter several alternatives has been tested and every additional of them accounted less problems. Even if the last version of the adapter is functional, it is still under development.

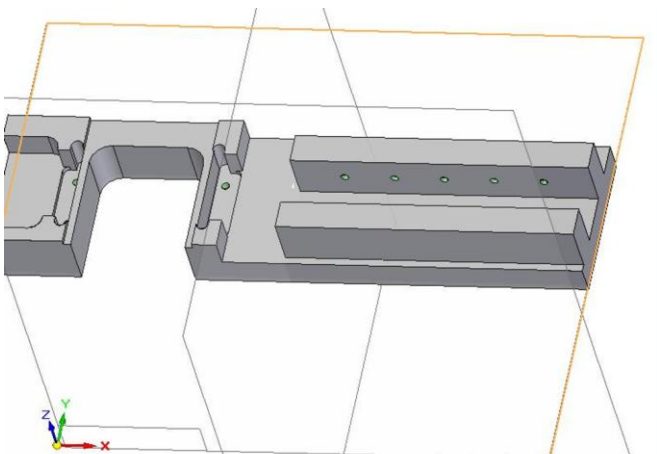


Fig. 2. Graphic model of micro-axial tomograph. This model shows body of micro-axial tomograph without stepper board and system of catching the capillary in front of the tomograph.

#### E. Glass capillary

For convenient analysis of cells and their consistent 3D reconstruction glass capillary was used. It is the most suitable

for spotted objects. Their application brought several disadvantages.

Very known phenomenon in the optic microscopy is optic deformation of objects, which is caused by shatter light ray by the passage of two regions with different optic density. This phenomenon is possible to compensate using immersion oils with convenient refractive index. Samples prepared in solute with convenient refractive index were sucked into the capillary and at the same time the entire capillary immerse into the same oil.

Other disadvantage of capillary is possible unevenness of its surface, which is generated during its production. Therefore it is necessary to produce them with microscopic precision. Experiments were carried out maximally with 100- multiple zoom, but even though unevenness of the capillary manifested.

Even the capillary is in certain extent flexible, manipulation with it is strenuous, because of its fragility.

During the experiments were used hand drawn capillaries from siliceous glasses. Disadvantage is that it is impossible to prepare capillaries with sufficiently equal precision. Problems come into being in the part of thick portions of capillary, so in regions where it changed thickness. Here was usually offset from straight way. This induced that capillary was deflected from visual fields by microscopy.

### III. IMAGE ACQUISITION AND ANALYSIS

#### A. Image acquisition strategies

Capillary was landed to adapter and fit in place in microscope. Subsequently it was adjusted that axis x of the adapter was the same as x axis of desk. It is necessary to minimize imprecision resulting from imprecision of place and unevenness of capillary.



Fig. 3. Glass capillary. One of capillary which was hand made and used. On the picture we can see unevenness of the capillary, which produces defocusing of microscope view and moving sample out of field of microscope view.

Collection consists of the following parts:

- sharpen to the top of capillary
- find the object of interest
- sharpen to the top of an object
- record of image information from current position
- rotate capillary and repeat the process to full of 360°
- move on the next region of capillary

Acquired data was saved as single picture with specific label. Label was characterized with x, y and z axis, actual angle of view and name.

### B. Image analysis

Data was processed in software Ellipse. Noise was removed and brightness was adjusted with histogram equalization. Finally all objects were extracted with local thresholding that separate objects if interest from background. Results were objects captured from different angles and saved.



Fig. 4. Starch grain. Sample of starch grain placed into the capillary acquired without using of immersion oil and without any treating. We used bifocal microscope and sample was lighted from down. In the picture shadows of objects are seen inserted in capillary and little deformation of form of capillary walls.

### C. Software

Stepper board was controlled with software specially modified to this engine in Visual Studio 2008 with C# language. It was debugged on Notebook with CPU DualCore 2GB, 3GB RAM, 64KB VGA on Windows XP Pro. The software displayed webcam view at real time and full control of the stepper board, and allowed automatic creation of images by certain angle of rotation.

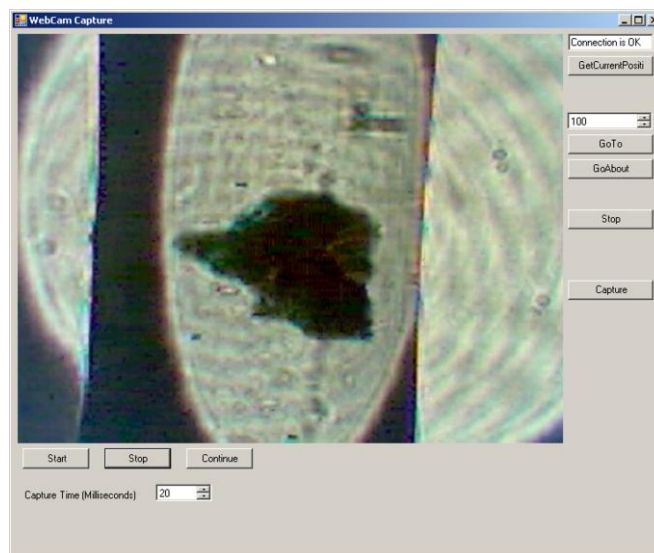


Fig. 5. Software GUI during sample acquisition.

## IV. CONCLUSION

The goals of the project were not obtained until now, because time to finalize was not still elapsed. The next step is to verify of the used devices and increase precision of capture and processing of data. We will be able reconstruct all captured object in 3D area in the final phase and therefore we will be able to observe objects with more details.

## ACKNOWLEDGMENT

This work was supported by Research and Development Support Agency under project APVV-0682-07.

## REFERENCES

- [1] P. Matula, M. Kozubek, F. Staier, M. Hausmann. "Precise 3D image alignment in micro-axial tomography", February 2003.
- [2] M. Kozubeka, M. Skalníková, P. Matula, E. Bártová, J. Rauch, F. Neuhaus, H. Eipel, M. Hausmann, "Automated microaxial tomography of cell nuclei after specific labelling by fluorescence in situ hybridisation", 2002.
- [3] R.J. Skaer, S. Whytock, "Interpretation of the three-dimensional structure of living nuclei by specimen tilt.", 1975.
- [4] P.J. Shaw, D.A. Agard, Y. Hiraoka, J.W. Sedat, "Tilted view reconstruction in optical microscopy: three-dimensional reconstruction of *Drosophila melanogaster* embryo nuclei.", 1989.
- [5] J. Bradl, B. Rinke, A. Esa, P. Edelmann, H. Krieger, B. Schneider, M. Hausmann, C. Cremer, "Comparative study of threedimensional localization accuracy in conventional, confocal laser scanning and axial tomographic fluorescence light microscopy.", 1996b.

# CoAlgebras and Object-Oriented Paradigm

<sup>1</sup>Marián JENČIK

<sup>1</sup>Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

<sup>1</sup>marian.jencik@mail.t-com.sk

**Abstract**—Object-Oriented Programming has become very common concept but only few people agree on what it means. In our paper we try to use coalgebras to explain objects and classes of object-oriented paradigm. Classes may occur as models of coalgebraic specifications and objects belonging to a class are elements of the state space of the class. We further describe differences between algebras and coalgebras and then their advantages for object-oriented paradigm.

**Keywords**—algebras, coalgebra, object, class, morphism

## I. INTRODUCTION

The object-orientated paradigm is derived from the convergence of other fundamental paradigms, and is reducible to the other paradigms as required by its application via a language or method. The flexibility provided by utilizing this paradigm can be best understood by examining the roots of the paradigm itself. The general idea of object-oriented paradigm is support for data abstraction with the ability to define and use new types. Data abstraction is one from thr other styles of programming and coalgebras is one of the method how to describe behaviour of these programs. Using of coalgebras to describe behaviour of programs we detailly explain in the next text.

## II. ALGEBRAS AND COALGEBRAS

Algebra and coalgebra are concepts used to describe some kinds of mathematical structures which are commonly met in mathematics and in computer science. The relationship between algebras and coalgebras appears clear only when their definition is formulated inside category theory: *Algebra* and *coalgebra* are dual concepts. This duality has been observed informally for long time, with algebras used to describe data types, and coalgebras used to describe systems (i.e., abstract machines).

An *algebra* is commonly described as a set plus some operations on it, and as such is usually used to formalize many kinds of data types in programming languages.

The general definition of an *algebra* is a map of the form  $T(\mathcal{X}) \rightarrow \mathcal{X}$ , for some functor  $T : \mathbf{Sets} \rightarrow \mathbf{Sets}$  on the category of sets and functions, where  $\mathcal{X}$  is state space, also called the carrier [1].

A *coalgebra* is a map  $\mathcal{X} \rightarrow T(\mathcal{X})$  in the reverse direction. Such a coalgebra consists of the same state space  $\mathcal{X}$  together with a transition function (or dynamics)  $\mathcal{X} \rightarrow T(\mathcal{X})$  acting on the state space [2].

If we have two algebras  $(c : T(\mathcal{U}) \rightarrow \mathcal{U})$  and  $(d : T(\mathcal{V}) \rightarrow \mathcal{V})$ , then we say that an algebra map  $c \rightarrow d$  is a homomorphism  $f : \mathcal{U} \rightarrow \mathcal{V}$  between the *carriers* which commute with the operations :  $f \circ c = d \circ T(f)$  as it is illustrated on Fig. 1.

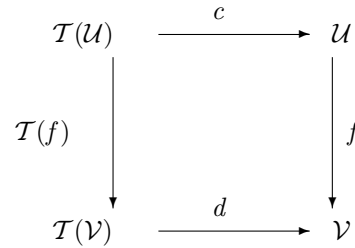


Fig. 1. Morphism between algebras

A *coalgebra map* consists of two coalgebras  $(c : \mathcal{U} \rightarrow T(\mathcal{U}))$  and  $(d : \mathcal{V} \rightarrow T(\mathcal{V}))$  is a homomorphism  $f : \mathcal{U} \rightarrow \mathcal{V}$  with operations :  $d \circ f = T(f) \circ c$  as it is illustrated on Fig. 2.

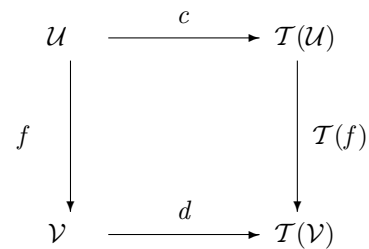


Fig. 2. Morphism between coalgebras

We can construct category  $\text{Alg}(T)$  consisting of algebras as objects and homomorphisms between them as category morphisms. Dually we can construct a category  $\text{CoAlg}(T)$  of coalgebras of  $T$ .

The essential difference between algebras and coalgebras is that the *algebras* have *constructors* (operations going into the underlying carrier set, which are used to build elements) where the *coalgebras* have *destructors* (operations going out of the carrier set, which allow us to observe certain behaviour)[3]. In other side *coalgebras* are looking at state space as black boxes to which have limited access via specified operations. It means, that behavior of objects can be observed via specified coalgebraic operations that we have in our disposal.

## III. OBJECT-ORIENTED PARADIGM

Object-oriented programming (OOP) is a technique for programming - a paradigm for writing *good* programs for a set of problems. The main idea of OOP is that data and procedures are represented by structures call objects. OOP consequently joins data model and procedural model in one



which are including in classical (non object-oriented) analysis and design.

The object-oriented paradigm focuses on the behavioral and structural characteristics of entities as complete units [4]. OOP uses two main entities : objects and classes.

*Objects* encapsulate structural characteristics known as attributes and behavioral characteristics known as operations. Attributes are representational constructs of structural characteristics of entities and determine the possible states of an object. Operations are representational constructs of behavioral characteristics of entities and determine the possible behaviors of an object. Objects have identities and are instances of classes. Fundamentally, objects are abstract entities that encapsulate state and behavior.

*Classes* are descriptions of objects with a common implementation. Classes are concerned with the implementation of uniform structural characteristics and behavioral characteristics. Fundamentally, classes are descriptions of objects with common attributes, operation implementations, semantics, associations, and interactions [4].

The object-oriented paradigm uses following pillars to increase productivity and consistency of programs [5].

- *Abstraction* involves the formulation of representations by focusing on similarities and differences among a set of entities to extract common features to define a single representation having those characteristics that are relevant to defining every element in the set.
- *Encapsulation* involves the packaging of representations by focusing on the hiding of details to facilitate abstraction.
- *Inheritance* involves the relating and reusing of existing representations to define new representations.
- *Polymorphism* involves the ability of new representations to be defined as variations of existing representations.

#### IV. COALGEBRAIC DESCRIPTION OF OBJECTS AND CLASSES

In this section we try to explain behaviour of object-oriented programs coalgebraically. Every object-oriented program consist of structures names objects and classes. In coalgebraically description are classes presented as coalgebraic models of class specifications and objects as inhabitants of the carrier set of class [2].

##### A. Objects

In object-oriented programming procedures are called methods [6]. Let have class  $(c : \mathcal{X} \rightarrow \mathcal{T}(\mathcal{X}))$ . An *object*  $x \in \mathcal{X}$  of the carrier set  $\mathcal{X}$  of the class is the method with the concept

$$c : \mathcal{X} \times I \rightarrow E + O \times \mathcal{X}$$

meaning that for each  $x \in \mathcal{X}$  and input  $i \in I$  either raises an exception E or yields an output O and a new state in  $\mathcal{X}$ .

##### B. Classes

A *class* in object-oriented programming is given as a set of the methods which can be written as

$$c_j : \mathcal{X} \rightarrow (E_j + O_j \times \mathcal{X})^{I_j} (1 \leq j \leq n)$$

The endofunctor  $\mathcal{T}$  associated with this signature of, say  $n$  such coalgebraic operations is

$$\mathcal{T}(\mathcal{X}) = \prod_{1 \leq j \leq n} (E_j + O_j \times \mathcal{X})^{I_j}$$

A *class specification* with functor  $\mathcal{T}$  consists of three elements :

- A carrier set  $\mathcal{X}$ , giving an interpretation of the state space
- A coalgebra  $(c : \mathcal{X} \rightarrow \mathcal{T}(\mathcal{X}))$  interpreting the methods as coalgebraic operations which are nothing else than a single function

$$\langle f_1 \dots f_n \rangle : \mathcal{X} \rightarrow \prod_{1 \leq j \leq n} Y_j$$

- An initial state  $x_0 \in \mathcal{X}$  which satisfies the condition in the creation section of the class specification

An implementation of *class specification* to be a system as it is illustrate on Fig. 3.

$$\mathcal{X} \xrightarrow{\langle c_1 \dots c_n \rangle} \mathcal{T}(\mathcal{X})$$

Fig. 3. System of implementation

If some output  $O_j$  is the empty set 0, then the associated method gets the concept  $\mathcal{X} \times I_j \rightarrow E_j$ , and may be called an *attribute*, since it yields an *observable element* in  $E_j$  and does not change the local state space [7].

The main reason why this view of classes as systems is attractive, is that it naturally takes into account that objects are encapsulated: The only way to access an object is via one of the methods. Therefore, there is for each class a notion of behavioural equivalence which expresses that two objects are equivalent if they cannot be distinguished by applying the methods to them [6]. This notion of behavioural equivalence coincides with the one given by the final system.

#### V. CONCLUSION

In our contribution we have mentioned some foundations of algebras, coalgebras and differences between them. We have presented our main ideas of object-oriented paradigm and coalgebraic description of objects and classes. The main strengths of using coalgebras are their simplicity and power in description of system behaviour what is the main subject of my future work.

#### ACKNOWLEDGMENT

This work was supported by VEGA Grant No.1/0175/08: Behavioral categorical models for complex program systems.

#### REFERENCES

- [1] M. Barr and C. Wells, *Category Theory for Computing Science*. Prentice Hall International, 1990.
- [2] B. Jacobs, "Objects and Classes Co-Algebraically," In: B. Freitag, C.B. Jones, C. Lengauer, and H.-J. Schek (eds) *Object-Oriented Programming with Parallelism and Persistence Kluwer Acad. Publ.*, pp. 83–103, 1996.
- [3] B. Jacobs and J. Rutten, "A Tutorial on (Co)Algebras and (Co)Induction," *EATCS Bulletin*, vol. 62, pp. 222–259, 1997.
- [4] S. S. Alhir, "The Object-oriented Paradigm," *O'Reilly Associates, Inc.*, October 1998.
- [5] V. Bono, "Type Systems for the Object-Oriented Paradigm," *PhD. Thesis, Universita di Torino*, 1996.
- [6] A. Kurz, "Coalgebras and Modal Logic," *Course Notes for ESSLLI 2001, Department of Philosophy, University of Helsinki*, October 2001.
- [7] K. S. Fisher, "Type systems for Object-Oriented Programming Languages," *PhD. Thesis, Stanford University*, 1996.

# Graph Cut Segmentation

Peter KARCH

Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

peter.karch@tuke.sk

**Abstract**—This paper deals with image segmentation based on Graph Cut method.

**Keywords**—Energy minimization, Graph Cut, image segmentation, maximum flow, minimum cut.

## I. INTRODUCTION

Segmentation is generally defined as the problem of partitioning an image into two or more segments: objects and background.

There are mainly three approaches to segmentation:

- automatic
- manual
- interactive

Manual segmentation is very laborious and extremely time consuming. Purely automatic segmentation is very challenging, due to ambiguities in the presence of multiple objects, weak edges, image noise, etc. Interactive segmentation is becoming more popular to alleviate the problems inherent to fully automatic segmentation technique that divides the image into two segments: object and background. A user imposes certain hard constraints for segmentation by specified certain pixels that are called seeds that absolutely have to be part of the object and certain pixels that absolutely have to be part of the background. Intuitively, these hard constraints provide information on that user intends to segment. The cost function is defined in terms of boundary and region properties of the segments. These properties can be perceived as soft constraints for segmentation. The details of this segmentation method are shown in Section II D. In next section will be explain the terminology for graph cuts and provide some background information. Section III provides an example of application graph cut algorithm using [4] implemented in Matlab for image segmentation.

## II. GRAPH CUT SEGMENTATION

Segmentation using this method belong between interactive methods which partitioning an image into two segments “objects” and “background”. This method is initialized by interactive or automated identification one or more points representing the object and one or more points representing background. These points are called seeds and they are represented hard constraints of segmentation. Other soft constraints are attributable to the border or regional information.

### A. Background on graphs

A flow network  $G(V, E)$  is defined as a fully connected,

directed graph where each edge  $(u, v) \in E$  has a positive capacity  $c(u, v) \geq 0$ . Then two special vertices in a flow network are designated as the source  $s$  and the sink  $t$  called terminals. A flow in  $G$  is real-valued function  $f: R^+ \cup 0 \rightarrow E$ , which each edge  $(u, v)$  assigned a nonnegative number  $f(u, v)$  so that satisfies the following three properties:

- capacity constraint:  
for all  $u, v \in V$ ,  $f(u, v) \leq c(u, v)$
- skew symmetry  
for all  $u, v \in V$ ,  $f(u, v) = -f(v, u)$
- flow conservation  
for all  $u \in (V - \{s, t\})$ ,  $\sum_{v \in V} f(u, v) = 0$

The value of a flow is defined as  $|f| = \sum_{v \in V} f(s, v)$ , and interpreted as the total flow from the source in  $G$ . A cut is a split of the nodes into two sets  $S$  and  $T$ , such that  $s$  is in  $S$  and  $t$  is in  $T$ . The capacity of a cut  $(S, T)$  is defined as

$$c(S, T) = \sum_{u \in S} \sum_{v \in T} c(u, v) \quad (1)$$

of purpose to find a cut with minimum capacity.

### B. Energy minimization

There are currently defined two basic models that define the minimization of energy  $E$ . They were selected basically on properties of labeled regions such as:

- Piecewise consistency
- Border discontinuities.

The first model is Potts Interaction Energy Model, defined as:

$$E(I) = \sum_{p \in P} |I_p - I_p^0| + \sum_{(p,q) \in N} K_{(p,q)} \cdot T(I_p \neq I_q) \quad (2)$$

where  $I = \{I_p | p \in P\}$  are the unknown true labels over the set of pixels  $P$  and  $I^0 = \{I_p^0 | p \in P\}$  are the observed labels corrupted by noise. The Potts interactions are specified by  $K_{(p,q)}$  what are the penalties for label discontinuities neighboring pixels. The function  $T(I_p \neq I_q)$  is an indicator function and it is 1 if condition inside the parenthesis is correct, and 0 if it is incorrect. Potts model is useful when the labels are piecewise constant with discontinuities at boundaries.

The second model is Linear Interaction Energy Model, defined as:

$$E(I) = \sum_{p \in P} |I_p - I_p^0| + \sum_{(p,q) \in P} A_{(p,q)} \cdot T(I_p \neq I_q) \quad (3)$$

where constant  $A_{(p,q)}$  describe the relative interaction importance between neighboring pixels. This model produces labels which are piecewise smooth but with discontinuities at

boundaries.

### C. Graph Cut

This method is one of the graph algorithms, where the main idea is to create valued, undirected graph  $G = (V, E)$  from image  $I$ . The graph is defined as a set of nodes or vertices  $V$ , which represents itself pixels in image, and a set of edges  $E$  represents any neighborhood relationship between the pixels and valued by nonnegative cost  $w_e$ . Then they are two special nodes named terminals  $S$  (source) and  $T$  (sink) that represents object and background labels.

Normally, there are two types of edges in the graph: *n-links* and *t-links*. *N-links* connect pairs of neighboring pixels or voxels. Thus they are represented a neighborhood system in the image. Cost of *n-links* corresponds to a penalty for discontinuity between the pixels. *T-links* connect pixels with terminals (labels). The cost of a *t-links* connecting a pixel and a terminal corresponds to the penalty for assigning the corresponding label to the pixel. This is normally derived from the data term (6) in energy (5).

A cut  $C$  is a subsets of edges  $C \subset E$  which when removed from  $G$  disjoint  $V$  into two subsets  $S$  and  $T = V - S$  such that  $s \in S$  and  $t \in T$ . Than cost of cut is defined as the sum of cost of this edges:

$$|C| = \sum_{e \in C} w_e \quad (4)$$

The minimum cut is the cut with smallest cost.

### D. Segmentation

The problem of segmenting an object from its background is interpreted as a binary labeling problem. An image is divided into set of  $I$ , which represented all pixels in the image and set of  $N$ , which represented standard 4-connected neighborhood system of pixel pairs  $\{p, q\} \in I$ , where  $q$  is neighborhood of pixel  $p$  on left, right, top or bottom from him.

Each pixel in image has assigned to the binary label  $L_k = \{O, B\}$ , where  $O$  represent object and  $B$  represent background, respectively. These two levels are identified by terminal nodes source  $S$  and sink  $T$ . The labeling vector  $L = (L_1, \dots, L_p, \dots, L_{|I|})$  defines a segmentation.

Then, the soft constraints that we impose on boundary and region properties of  $L$  are described by the cost function

$$E(L) = \lambda \cdot R(L) + B(L) \quad (5)$$

where

$$R(L) = \sum_{p \in I} R_p(L_p) \quad (\text{regional term}) \quad (6)$$

$$B(L) = \sum_{(p,q) \in I} B_{(p,q)} \delta(L_p, L_q) \quad (\text{boundary term}) \quad (7)$$

and

$$\delta(L_p, L_q) = \begin{cases} 1 & \text{if } L_p \neq L_q \\ 0 & \text{if } L_p = L_q \end{cases}$$

The coefficient  $\lambda \geq 0$  in (1) defines the importance of the regional properties term  $R(L)$  versus the boundary properties term  $B(L)$ . The regional term assumes that the individual penalties for assigning pixel  $p$  to object and background, correspondingly  $R_p(O)$  and  $R_p(B)$ , are given. This  $R_p(O/B)$  reflect how the intensity of pixel  $p$  fits into given intensity model (histogram) of the object and background

$$R_p(O) = -\ln P(I_p|O) \quad (8)$$

$$R_p(B) = -\ln P(I_p|B) \quad (9)$$

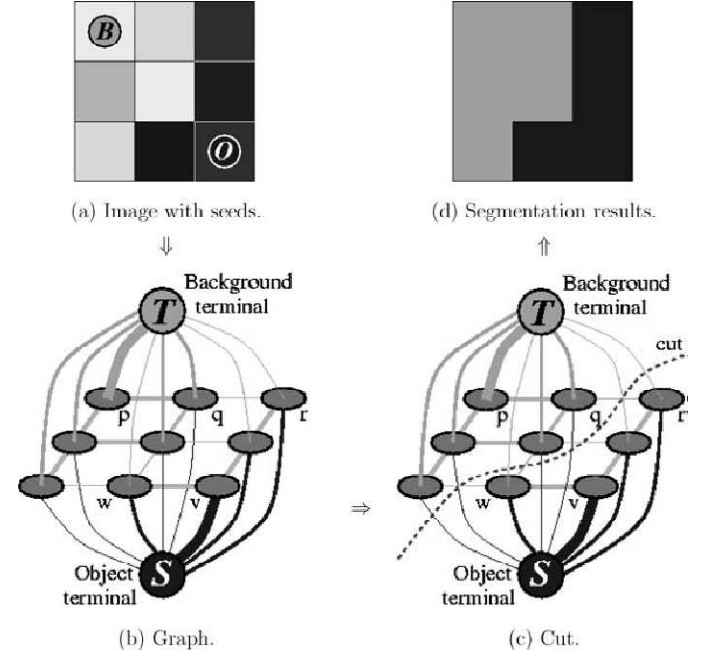
The boundary term  $B(L)$  comprises boundary properties of segmentation and determined smoothness on boundary.

Coefficient  $B_{(p,q)} \geq 0$  can be interpreted as penalty for discontinuities between  $p$  and  $q$ . Normally,  $B_{(p,q)}$  is small when pixels  $p$  and  $q$  are very different and  $B_{(p,q)}$  is large when they are similar. The cost  $B_{(p,q)}$  can be determined

$$B_{(p,q)} \propto \exp\left(-\frac{(I_p - I_q)^2}{2\sigma^2}\right) \cdot \frac{1}{\text{dist}(p,q)} \quad (10)$$

where  $\text{dist}(p,q)$  is the Euclidean pixel distance and  $\sigma$  is user defined parameter.

A simple 2D segmentation example is shown in Fig. 1.



**Fig. 1 A simple 2D segmentation example for 3 x 3 image. (a) The seeds are  $O = \{v\}$  corresponding to object and  $B = \{p\}$  corresponding to background. (b) Graph. (c) Graph Cut. (d) Segmentation result.**

## III. EXPERIMENTS

This implementation in experiments uses a new max-flow algorithm from [2] implemented in [4]. Segmentation was performed on grayscale image *chromozomes.jpg*[10]. Actual implementation [4] doesn't allow a user to enter seeds. They are obtained using the k-means function which assign each pixel to one of two labels. Function GraphCut is looking for segmentation by optimization of the function:

$$E(L) = R(L) + B(L)$$

where

$$R(L) = \sum_{p \in I} D_p$$

is the distance of the each of pixels of each label (calculated using k-means). For simplicity is used Euclidean norm

$$D_p = \sqrt{(L_p^2 - l_k^2)} \quad (l_k - \text{value of } O/B \text{ label}).$$

$$B(L) = \sum_{(p,q) \in I} S(L_p, L_q)$$

is regularize the resulting segmentation and it's defined by matrix  $S(L_p, L_q)$ .  $B(L)$  penalize isolated pixels in segmentation and prefers larger areas. Each item of the matrix represents the jump cost between  $L_p$  and  $L_q$ . If they are consistently  $L_p = L_q$  than cost is 0 other it is  $\delta$ .  $S(L_p, L_q) = \begin{cases} 0, & L_p = L_q \\ \delta, & L_p \neq L_q \end{cases}$ .

Increasing this constant reinforces the neighbor conditions in

the segmentation.

```
Some GraphCut function usage:
handle = GraphCut('open',Dc,Sc);
[gch l] = GraphCut('expand',handle);
handle = GraphCut('close',handle);
```

handle - A handle of the constructed graph.  
 l - a width\*height array of type int32, containing a label per pixel  
 'open' - Create a new graph object.  
 'expand' - Perform labels expansion  
 'close' - Close the graph and release allocated resources  
 DataCost(Dc) – equals to the cost for assigning labels to pixel  
 SmoothnessCost(Sc) - matrix where Sc(l1, l2) is the cost of assigning neighboring pixels with label1 (O) and label2 (B).  
 Vc,Hc - optional arrays defining spatially varying smoothness cost

An example in Fig.2 shown what effect have the increasing  $\delta$  constants on segmentation result.

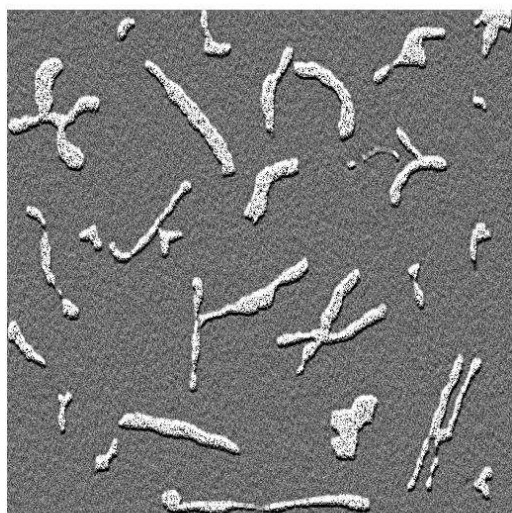


Fig. 2a Original image

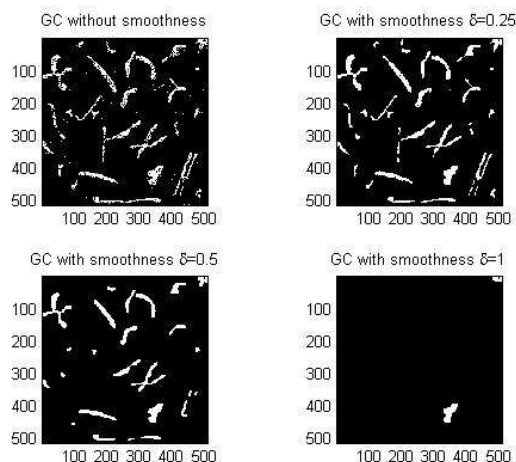


Fig. 2b Graph Cut without smoothness and with smoothness of various  $\delta$

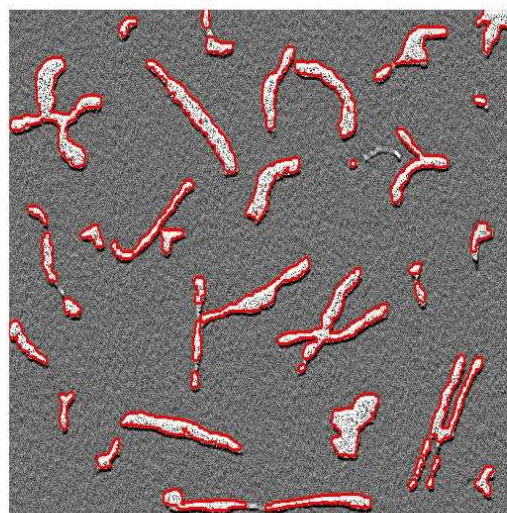


Fig. 2c Graph Cut segmentation with value of  $\delta = 0.25$

The data and smoothness terms by themselves provide a good segmentation. However, the results can be further improved if edge information is also taken into account, to encourage pixel label changes across edges and discourage them otherwise. The edge information (separately for horizontal and vertical directions) was obtained by applying a smoothed Sobel filter. The horizontal and vertical costs are then passed to GraphCut('open',...) as additional parameters  $V_c, H_c$ .

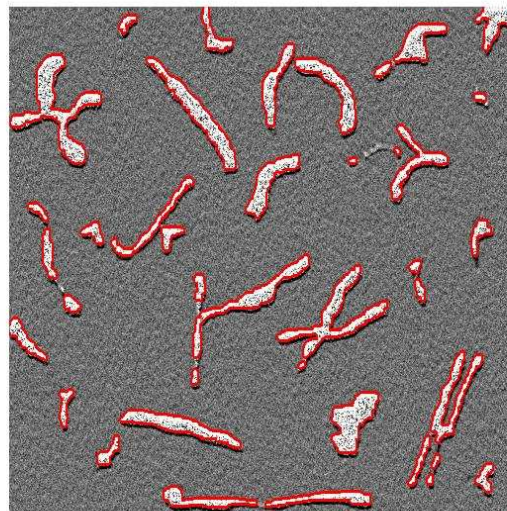


Fig. 2d Graph Cut segmentation with value of  $\delta = 1$  and edge term

Matlab Wrapper for Graph Cut [4] is compiled and implemented in Matlab R2007a. All experiments results are given with CPU Core2Duo 2.13GHz, 3GB RAM and 256MB VGA on Windows XP Pro.

#### IV. CONCLUSION

Implementation [4] is not final whereas terminals identification is not interactive and labels are obtained using k-means into matlab.

The next step in others experiments is to create GUI (graphical user interface) and implementation Graph Cut segmentation under Visual C++.

#### ACKNOWLEDGMENT

This work was supported by grant APVV-0682-07.

#### REFERENCES

- [1] Y. Boykov, V. Kolmogorov, "An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization In Vision", IEEE transactions on Pattern Analysis and Machine Intelligence, vol. 26, no.9, pp 1124-1137, September 2004.
- [2] Y. Boykov, O. Veksler, R. Zabith, "Efficient Approximate Energy Minimization via Graph Cuts", IEEE transactions on Pattern Analysis and Machine Intelligence, vol. 20, no.12, pp 1222-1239, November 2001.
- [3] V. Kolmogorov, R. Zabih, "What Energy Functions can be Minimized via Graph Cuts?", IEEE transactions on Pattern Analysis and Machine Intelligence, vol. 26, no.2, pp 147-159, February 2004.
- [4] S. Bagon, "Matlab Wrapper for Graph Cut", December 2006. Online: <http://www.wisdom.weizmann.ac.il/~bagon/matlab.html>
- [5] Y. Boykov, G. Funka-Lea, "Graph Cut and Efficient N-D Image segmentation", International Journal of Computer Vision 70(2), 109-131, 2006.
- [6] T. Collins, "Graph Cut Matching In Computer Vision", February 2004.
- [7] S. Sinha, "Graph Cut algorithms in Vision, Graphics and Machine Learning, An Integrative Paper", UNC Chapel Hill, 2004.
- [8] T. H. Cormen, C.E. Leirson, R.L. Rivest, C. Stein "Introduction to Algorithms", McGraw – Hill, 1990.
- [9] M. Sonka, V. Hlavac, R. Boyle, "Image Processing, Analysis, and Machine Vision", 2007.
- [10] [Online]. Available: <http://visionbook.felk.cvut.cz/>

# Content management systems using ontology

<sup>1</sup>Ján KAŽIMÍR

<sup>1</sup>Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

<sup>1</sup>jan.kazimir@tuke.sk

**Abstract**—in this paper we shall discuss using of ontology to provide functionality and base for Content management system. Then we shall look on advantages and disadvantages of ontologies used for Content management system.

**Keywords**—content management, content management system, digital content, ontology.

## I. INTRODUCTION

For a lot of really big companies and corporations is vital to maintain knowledge in some form of documents and so to manage whole live cycle of those documents. The existence of those documents is only half of a problem. Second half is to publish correct document the correct way. The answer to this common problem is Content management system. Those are really hard to create and maintain. One way to ease Content management systems can be by applying of ontology methodology.

## II. ONTOLOGY

Word *ontology* is nothing new and there are two different definitions for ontology. First is from philosophy. Second definition of ontology, relevant for this paper, is look on ontology from computer science side.

### *Definition of ontology*

In the context of computer sciences, an ontology defines a set of representational primitives with which to model a domain of knowledge or discourse. The representational primitives are typically classes (or sets), attributes (or properties), and relationships (or relations among class members). The definitions of the representational primitives include information about their meaning and constraints on their logically consistent application. In the context of database systems, ontology can be viewed as a level of abstraction of data models, analogous to hierarchical and relational models, but intended for modeling knowledge about individuals, their attributes, and their relationships to other individuals [4].

Ontologies are typically specified in languages that allow abstraction away from data structures and implementation strategies; in practice, the languages of ontologies are closer in expressive power to first-order logic than languages used to model databases. For this reason, ontologies are said to be at the "semantic" level, whereas database schema are models of

data at the "logical" or "physical" level. Due to their independence from lower level data models, ontologies are used for integrating heterogeneous databases, enabling interoperability among disparate systems, and specifying interfaces to independent, knowledge-based services. In the technology stack of the Semantic Web standards [1], ontologies are called out as an explicit layer. There are now standard languages and a variety of commercial and open source tools for creating and working with ontologies.

## III. DIGITAL CONTENT AND CONTENT MANAGEMENT

If we are to talk about content management systems, we have to know, what *content* we are talking about. To be more specific, in our case is important to know the meaning of phrase "*digital content*". In global there are two approaches of understanding this phrase.

### A. *Digital content – narrow sense*

In narrow sense, the digital content include only those data or parts of data that we understand as any type of information. So in this case the digital content means images, texts, sounds, animations, videos and etc. We will by using this meaning of phrase "digital content".

### B. *Digital content –wide sense*

In wide sense, the digital content is everything that is digitally represented. That means that in this case the digital content includes not only information, but also programs that represent this information to us.

For this paper we consider to use an organization of digital content to documents. That also lead in to call those fragments of digital content the documents.

Now that we know what content we are managing, we can proceed to definition of *Content Management*.

Content Management can be defined as a framework to generate, to organize and prepare, to distribute and to create possibilities of using and processing digital content that can be located on the Internet, Intranet or corporation-wide systems. Attention is drawn on actuality, consistence and accessibility of the content.

This framework is integrated to an internet platform to manage Web content, or it can be running on local platform to manage digital content accessible only on this local platform.

#### IV. CONTENT MANAGEMENT SYSTEM

Content Management System (CMS) is a computer application that creates, edit, manage, search and publish various kinds of digital content. So it is concrete application used for content management. CMS is inferred from evolution of *live cycle of digital content* that this system has to manage.

CMS can be divided to three sub-applications as we can see on figure Fig. 1:

1. *Content Management Application (CMA)* – an application that is directly evolved in document live cycle (creating, editing, deleting content)
2. *Meta Management Application (MMA)* – an application that manage meta-data (knowledge and properties) of concrete document.
3. *Content Delivery Application (CDA)* – an application that interprets and publish documents.

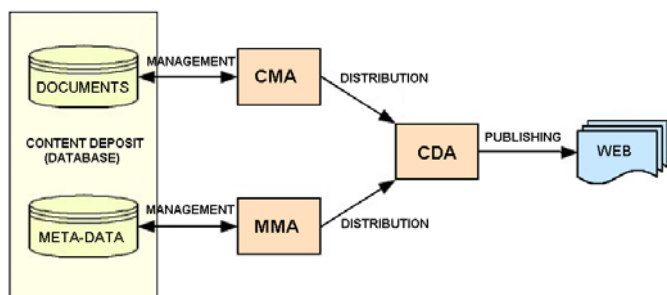


Fig. 1. Structure of Content management system

CMA is not necessarily one application. It can be a group of applications that provide the functionality of CMA. There are many already developed applications that we can use for this purpose.

As a CDA is mostly used an internet publishing application such as Internet Information Services or kind of an Apache software and there add-ons.

The most important part of CMS is MMA. All decisions are based on meta-data of document. CMS are useful on large depositories of documents. When we take these large depositories as complex system, it is more than likely to make a model of it. If we are able to make model from documents, then it is relatively easy to manage even big amount of documents.

#### V. ADVANTAGES AND DISADVANTAGES OF USE AN ONTOLOGY

Because the modeling helps us with the managing task, let us to consider using a concept of ontology to create MMA. It won't be real ontology; we will use only mechanisms used by ontologies. Those mechanisms are ontology languages, for example OWL, and inference mechanisms that are used to reason in ontologies. That gives us upper hand in management process with already existing theory and already existing tools. The biggest advantages of using ontology mechanisms are:

1. Existence of theory about processes in ontology
2. Existence of tools maintaining ontologies
3. Ontology is already standardized (languages, processes, etc.)
4. Easy sharing knowledge about domain (documents)

But standardization of CMS is also disadvantage, because it loses singularity. With ontology the CMS loses flexibility of making it different.

#### VI. CONCLUSION

The weights of advantages are more than disadvantages, especially if we talk about large deposits of documents. Ontology can help us with making large scale CMS that will be easy to transfer whole dataset to other platform.

#### REFERENCES

- [1] Berners-Lee, T. - Hendler, J. - Lassila, O.: „*The Semantic Web*“, Scientific American, May 2001. Also <http://www.w3.org/2001/sw>
- [2] Boiko, B.: „*Content Management Bible*“, John Wiley and Sons, 2004.
- [3] Gruber, T. R.: „*A Translation Approach to Portable Ontology Specifications*“, Knowledge Acquisition, 1993. See also <http://www-ksl.stanford.edu/kst/what-is-ontology.html>
- [4] Gruber, T. R.: „*Encyclopedia of Database Systems*“, Springer-Verlag, 2008.
- [5] Takáč, L.: „*Kybernetické modelovanie procesov spravovania systémov*“, dizertačná práca, Fakulta elektrotechniky a informatiky TU v Košiciach, 2008.

# P2P proxy/cache

*Ivan Klimek*

Computer Networks Laboratory, Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

Ivan.Klimek@cni.tuke.sk

**Abstract**—Internet as we know it nowadays is full of suboptimal behavior, widely used protocols are not optimized for factors not concerning the end users. This factors ultimately affect the whole Internet, a great example of such is the Peer-to-Peer (P2P) traffic problem. P2P traffic currently represents 50-90 percent [1][2][7] of whole Internet traffic, with the expectation to grow by 400+ percent in the next 5 years [2]. Most Internet Service Providers (ISP) try to minimize this traffic using some form of restrictive counter measures, like traffic shaping, Fair Use Policies (FUP) or similar techniques. This approaches restrict the end user experience, limit the possibilities in which they are able to use their Internet connection just to save ISP resources or delay the need to invest into new infrastructure, thus artificially slowing down the growth of Internet which in fact as this study will show is not necessary.

**Keywords**—caching, network optimization, P2P, redundancy

## I. PEER-TO-PEER PROTOCOLS

P2P networks are based on the idea of decentralization, sharing of resources across large number of hosts/users. This decentralization enables the network to perform in ways that cannot be compared to any other file/content providing technology. In all of the parameters as total download amounts, speeds, availability, scalability, costs - P2P superiority is unmatched. There are many P2P implementations currently available; the most used is the BitTorrent protocol with 60-90 percent of total P2P traffic [1]. BitTorrent is a very flexible protocol that can be used to deliver any kind of content. For example it is the ideal way for distributing content like Video-on-demand (VOD) and/or other high bandwidth applications. Because of the mentioned we will focus on BitTorrent in this study. The main motivation is that the clear benefits and possibilities of this technology can not to be fully exploited without prior solving the negative effects of decentralization. The biggest issue is that the data flows are mostly not controlled, that means the whole network communication is not effective and creates a lot overhead. For example the total P2P traffic is by more than 75 percent redundant [3].

## II. BITTORRENT ARCHITECTURE

The BitTorrent protocol became over the years of its existence pretty complicated, in this paper we won't try to describe the full architecture with all the extensions that currently exist. We will focus on the key element of this technology and that is the client-tracker communication. Tracker is a dedicated server that coordinates the clients; it

does not hold any content itself. Clients download the metafile (.torrent) from the tracker resp. from other sources, open it via their BitTorrent client software which then reads the information contained in the metafile as torrent information (date created, comment etc.) file information (name, length, hash) and tracker information (hostname, port). The client software then connects to the specified tracker and requests the torrent content using a hash generated from the information contained in the metafile, this hash is called infohash and is used in the whole BitTorrent protocol to identify the torrent's content. Tracker replies with a list of other clients that are downloading the same content (leechers) or already finished the download but are still uploading (seeders). The client software then starts the negotiation with other clients, but keeps updating the tracker in regular intervals. The tracker is updated when the download is completed too. The updates consist of information like: action (started, stopped, completed), bytes received, bytes left, bytes uploaded etc.

## III. PEER-TO-PEER PROXY/CACHE

By evaluating the architecture of the BitTorrent protocol, we were able to detect an attack vector using which it is possible to perform a Man-in-the-Middle attack on the client-tracker communication. By intercepting the client requests and tracker responses, we were able to create a transparent BitTorrent proxy/cache server. This device gives us the full control over the user downloads in the network segment behind the proxy. It is possible to collect detailed statistical information on all downloads and based on this select content to be cached and then transparently served to other users requesting the same content thus eliminating redundancy of data being downloaded.

Benefits of such approach:

- absolutely no uplink from clients in the given segment
- no redundant data outside of segment, massively reducing amounts of data downloaded
- if content is already in cache, clients download the content at full speed of their Internet connection (link) from the very first byte to the end, this is in contrast to classic P2P behavior where downloads start very slow and then by negotiating with more users gain speed
- if content is not yet cached and there is a high possibility more users will be interested in the same content the proxy actively supports the client download thus greatly enhances the client download speeds while parallel caching the content



- because of the full control over client downloads it is possible to optimize the network traffic
- specific content can be kept accessible longer than on classical P2P networks

Because the device is completely transparent it can be placed on different layers of the ISP network creating a hierarchical structure that maximizes mentioned benefits even more. Using proxies a hybrid P2P network can be created, hybrid because the content is no longer completely decentralized, the data flows are no longer uncontrolled and the negative side-effects of P2P are solved. The final effect of proxying depends on fine tuning the "all data - stored data" ratio. More TB disk space will enable more data being served from local cache. Even that the clients won't upload any data (effectively freeing the last mile uplink) the P2P network from global perspective is not affected as the cache itself is keeping uploading the content. It is important to mention that because of the transparency of this solution neither the client software nor any other part of the currently used technology has to be changed.

This technology can be used not only in standard wired or wireless ISP's networks but it also opens new possibilities for satellite Internet providers. Great majority of Earth's population has no access to broadband Internet, the latest generation of satellites is able to provide download speeds of up to 155Mbps with 6Mbps upload per user with relatively small dish size diameter of 45 centimeters [4]. The problem with this kind of Internet connection is the extremely slow latency generally about 500-900ms [5] (for comparison dial-up has latency around 150-200ms). The only way around this problem is to use on-orbit caching, thus reducing the latency for content provided from the cache to theoretical limit of 233ms (for the orbit height of 70,000 km). As shown with the use of extensive multi-protocol caching it would be possible to reduce the latency to dial-up levels for almost all non-real time traffic.

#### A. Intercepting Tracker HTTP/HTTPS protocol

Tracker requests are plain HTTP requests with a specific message format; this could be called the completely weakest point of the whole technology. For example some traffic filtering techniques use tracker responses which contain a clear text list of peers with their specific IP addresses and ports to deny just the Peer wire protocol communication (inter-client communication). As a reaction, some trackers started to encrypt this part of the response. Our proxy attacks the requests, it recognizes them and intercepts them and then the actual man-in-the-middle (MiTM) attack begins. Most trackers still use HTTP or at least support it for compatibility reasons, however we are able to intercept HTTPS in most cases too. This is because many trackers use self-signed certificates which are vulnerable to MiTM, when this is not the case then MiTM on HTTPS can be still successful because majority of BitTorrent clients does not verify the server certificate. Further, it is a de-facto standard to use always more than one tracker - multiplying the chance of interception. The only case in which this interception mechanism would not work would require that only one tracker is specified and that uses a CA signed certificate (or all specified trackers use HTTPS and have a CA signed certificate) simultaneously the BitTorrent client verifies the

certificate and then actively refuses the connection because it detects a MiTM attack. As of the date of writing this paper the author is not aware of any BitTorrent client that would act as mentioned.

#### B. The logic behind

The next phase after we intercepted the request is to decide if we already have the content cached, if yes, simply respond to the client as the original tracker providing the IP address of our proxy as the only peer who has the content, and the port number on which the content can be downloaded. If we do not have the content, we have to decide if it is "interesting" content i.e. it is viable to download it. This can be decided using the number of requests from the network served, for example if there will be more than x requests on this specific content then we cache it. Or we can use public search engines, look how many people already downloaded it, if it is popular in the last x hours then it is highly probable there will be more than one request for that file from our segment and it is a good idea to cache it. While we can't serve the requested content to the client, we forward the original tracker's respond to him. If we decided that it is an interesting content we will start to download it immediately, the client which originally requested the torrent is still downloading it and in regular intervals updating the tracker. In the next update he will get except of the "legal" peers the proxy in the list of peers too. The proxy will download the content faster than the client because of several factors:

- it is on a faster link
- it can accept incoming connections (firewalled clients can still download but their speeds are lower because their ports are blocked so they cannot accept incoming connections from other clients)
- it updates its list of trackers using all possible trackers that can be found on search engines having the content to be downloaded registered, thus it isn't limited only to the trackers specified in the metafile as the client; more trackers = more peers = more speed (it is searching for all trackers with the content's infohash registered)

The content is being provided to the client in parallel to downloading it thus massively enhancing the client's download speed.

#### C. Finding content

We developed two methods of caching content without the need to dump or "record" any network communication. First, the default one is based on BitTorrent search engines. The client sends the infohash parameter in its requests, we use it to search for that specific torrent metafile and then download it into the cache when needed. Using this method, it is possible to find most of the torrent metafiles, because the popular trackers are all well indexed. However, not everything can be found, for example if the content is served by some private or small community tracker. When that happens, the backup method is used. This method is by far more complicated than the default one, but is able to reconstruct all the information needed to download the torrent content just from the client requests and some protocol hacking. The problem is that the user starts the download based on information from the metafile; we do not have this metafile so if we want to download the content we need to get the missing information

somehow. The most important thing is to get the piece length, without this parameter it would be impossible to reconstruct the received data. Luckily, from information sent by other peers via the Peer wire protocol it is easy to calculate the piece length parameter from the bitfield length. Combination of this two methods results in almost 100 percent probability of being able to download all requested torrents to the local cache.

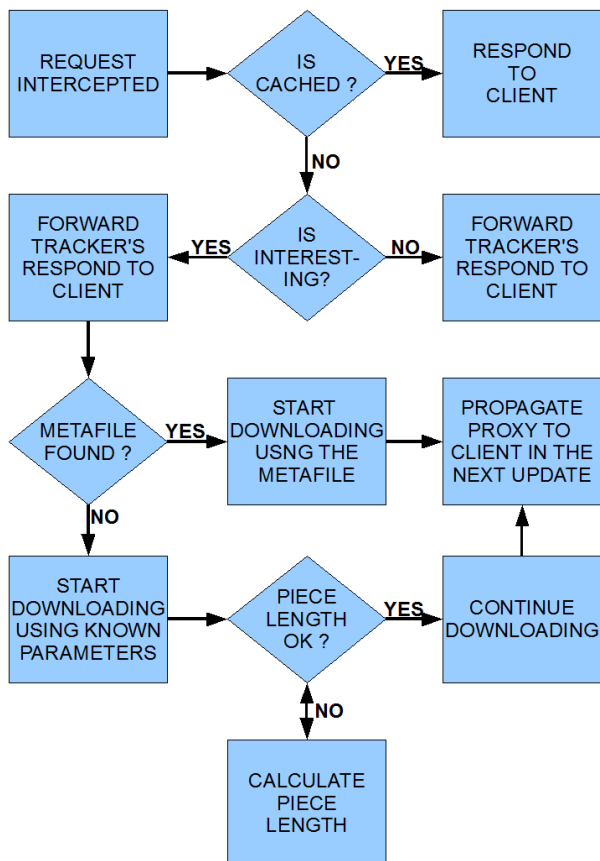


Figure 1: Simplified application logic

D. Full control

Because all BitTorrent downloads are being monitored by watching Tracker HTTP/HTTPS protocol, it is possible to alter the traffic in ways that suit the current situation/network technology/topology used. For example on wireless connections P2P represents a bigger problem than on wired because of the shared spectrum. Few P2P downloads can burden the whole sector, making it very hard to keep the QoS for all users on acceptable levels. As all P2P traffic is being controlled, it is possible to add a layer of virtual time division multiplexing, that manages downloads in a way that every time only one controlled data flow per segment exists. All clients recognize their downloads as active but in fact, they are getting data only few seconds per time period.

Benefits of such approach:

- P2P does not need to be restricted on wireless Internet connections
- P2P does not burden the sector excessively
- data speeds can be throttled dynamically based on network load

E. Software architecture

The Tracker HTTP requests are being intercepted by iptables [16] string matching rules. The HTTPS requests can be intercepted only by iptables "known tracker IP addresses" matching because before the initialization of the MiTM attack it is not possible to see the actual request. The iptables rules change the destination to localhost where our proxy is listening for it. Communication with the clients is done using the netcat tool [14] (with stunnel for HTTPS [15]). After the request is intercepted, it is parsed and the application logic as described earlier is performed. For downloading the content modified torrent is used (minimalistic console BitTorrent client software). Every instance of ctorrent is running on a separate port, we build a table describing on which port what content is being provided. Ctorrent was modified to be able to download the content of a torrent without the metafile (this approach is needed only when the metafile can't be found using BitTorrent search engines). It needs only the infohash, full length of the content and the tracker's address (one or more) parameters, all of this information can be parsed from the client's request. Ctorrent then starts downloading the torrent with default piece length, it listens to Peer wire protocol for the bitfields and using a special algorithm verifies if the default piece length is the right one. If not, then the download is restarted with the correct value.

The piece length calculation algorithm (C code representation):

```

size_t GetRealPieceLength(size_t bitfieldNBytes)
{
    size_t pieceLength = argm_file_size / (bitfieldNBytes * 8);
    return (log2(pieceLength) > int(log2(pieceLength))) ?
        2^(int(log2(pieceLength))+1) : 2^(int(log2(pieceLength)));
}
  
```

where:

bitfieldNBytes is the length of the bitfield

argm\_file\_size is the full content size

pieceLength is what we need to calculate

According to the BitTorrent protocol specification: "A bitfield of the wrong length is considered an error. Clients should drop the connection if they receive bitfields that are not of the correct size, or if the bitfield has any of the spare bits set." [6] that is why we can be almost sure if we see few same bitfield lengths from different sources it is the correct length.

The Bitfield identifies which block has the client already downloaded. Piece length is usually power of 2 [6], that is why we simply look for the nearest bigger power of 2 to the division of full content length by number of bits in the bitfield. The problem is that the spare bits on end of the bitfield are set to zero, so its length does not need to be always the correct argument. That means the maximal error can be 7 bits, a very small probability exists our piece length calculation could provide a bad result. Anyway, this is solved later in ctorent code, where specific errors can be produced only by the piece length miscalculation; the piece length is then automatically divided by two. This should provide correct setting of the piece length parameter, all the time. Another problem with downloading torrents without the initial metafile is the process of checking for bad downloaded pieces

as we do not have the SHA hashes to check the downloaded data with. This can be solved by another small modification of ctorrent, as we are the only peer the client can download the content from (if it is already cached), when we provide him a bad downloaded piece he will ask for it again and again. This behavior can be easily recognized and the identified piece re-downloaded from the network and provided to the client.

#### F. HW requirements

The presented solution does not need to be run on special HW, its requirements can be compared to normal file server HW for given bandwidth, although minor modifications consequent from the specifics of the protocol would greatly enhance its performance. The most heavily loaded component of a proxy is always the storage subsystem, but because the content will be mostly loaded just once into the cache and then multiple times read the ideal approach would be to combine cheap high-capacity storage (using classical disks) with extremely fast and extremely low access time read buffer (using SSD technology). The downloading of content is a slow process, everything is being cached onto the high-capacity storage, from that when needed copied onto the read buffer - SSD disk, all requests on that content are served from the buffer. The read buffer is dynamically cleaned so that only the currently most requested content is stored on it.

#### IV. COMPARISON WITH SIMILAR TECHNOLOGIES

There are four main competitors, all have the same goal, to solve the P2P traffic problem but use different approaches.

*P4P* - as described on its homepage: "*P4P is a framework that can be used to enable Internet service providers (ISPs) and peer-to-peer (P2P) software distributors to work jointly and cooperatively.*" [7] This means a special dedicated server in the ISP infrastructure coordinates the P2P data flows; it also needs special client software.

*pCache* - academic open-source project, aims to create a P2P cache too. Uses different approach, supports not only BitTorrent but Gnutella too. It is a more complicated solution because it needs to monitor all P2P data, resulting in higher HW requirements. [8][9]

*OverCache P2P Caching and Delivery Platform* - short description from pCache authors: "*Oversis's MSP platform realizes multi-service caching for P2P and other applications. An MSP device actively participates in P2P networks. That is, MSP acts as a ultra-peer that only serve peers within the deployed ISP. We believe this approach negatively impacts fairness in many P2P networks, such as BitTorrent, which employ algorithms to eliminate free-rider problem. In fact, no peers in ISPs with Oversis's MSP deployed will ever upload anymore, because they expect to get the data free from the MSP platform. Once number of free-riders increases, the P2P network performance degrades, which in turns affects P2P users all over the world.*" [8] Plus it is not a transparent solution.

*PeerApp UltraBand Family* - supports transparent caching of P2P traffic. Supported protocols are BitTorrent, Gnutella, eDonkey, and FastTrack. Uses Layer 7 protocol recognition (packet inspection) and according to that Layer 7 redirection to the application logic where it simultaneously to the client downloads acquires the file by dumping/cloning the

transmitted data. Layer 7 protocol recognition is mostly being avoided on firewalls because of great performance impacts, thus we believe the PeerApp solution performance will suffer from it. [10]

The complexity of our approach compared to caching solutions that need to monitor all BitTorrent packets can be demonstrated on the ratio between the Tracker HTTP/HTTPS and the Peer wire protocols. On an example download with very small content size of 4.4 MB this ratio is 4 to 5988 i.e. 1 to 1497! In larger content size downloads this ratio would be even bigger although there would be some regular tracker updates send. In most situations, our P2P proxy/cache will need to check/process only one packet per download - the initialization "event=started" tracker request. In the worst case scenario, except of this one packet it will need to process the "event=completed" and the tracker update messages which still represents only few packets. The piece length checking should not be calculated in the complexity because it does not create any overhead as it is just a parameter check added into the actual content download process.

The rest of the proxy operation consists just of downloading and providing content that means downloading a torrent and keeping seeding it until the cache capacity allows it i.e. it isn't deleted to provide space for a currently more popular content. This is very similar to the Oversis's approach. Just with the exception that it is transparent for the clients, but acts as a peer for the rest of the Internet to keep the content globally available. This is for change similar to PeerApp technology except it does not need to use Layer 7 redirection. The transparent approach maximizes the caching effectiveness because it simply said "hacks" into the client download process and is not affected by the peer propagation/selection of the tracker or client's settings/preference. The achieved 100 percent success rate of interception only demonstrates this.

As of the character of the presented solution, it is impossible to provide universal statistical results; it can simply work or not work. The only statistical results that could be provided are associated with the popularity distribution, but because this topic has been researched by numerous previous studies [9][11][12][13] we won't focus on it.

#### V. CONCLUSION

The presented solution by exploiting a found attack vector in the currently most popular and perspective P2P protocol - BitTorrent, enables the ISPs to stop fighting P2P and instead cooperate with the community on creating a more effective hybrid P2P network model. The P2P proxy/cache is an extremely lightweight and simple win-win solution when comparing with other similar approaches with at least the same or even better performance. This technology has been already tested in several test scenarios reflecting real life situations, always with great results. We are completely confident about the usability of this device and are very close to installing it into our university campus network.

## ACKNOWLEDGMENT

*The author would like to thank: Associate Professor Frantisek Jakab PhD., Tomas Korenko, Bc. Marian Keltika and Martin Chalupka.*

## REFERENCES

- [1] H. Schulze, K. Mochalski. Internet Study 2007, ipoque, November 2007.
- [2] MultiMedia Intelligence, “P2P Traffic to Grow Almost 400% over the Next 5 Years, as Legitimate P2P Applications Become a Meaningful Segment”, 2008, [Online; accessed 23-November-2008], [Online], Available: [www.multimediantelligence.com](http://www.multimediantelligence.com)
- [3] PeerApp, “UltraBand Family overview”, 2007, [Online; accessed 23-November-2008], [Online], Available: <http://www.peerapp.com/products-ultraband.aspx>
- [4] JAXA, “Overview of the KIZUNA (WINDS)”, 2008, [Online; accessed 23-November-2008], [Online], Available: [http://www.jaxa.jp/countdown/f14/overview/kizuna\\_e.html](http://www.jaxa.jp/countdown/f14/overview/kizuna_e.html)
- [5] Wikipedia, “Satellite Internet access”, 2008, [Online; accessed 23-November-2008], [Online], Available: [http://en.wikipedia.org/wiki/Satellite\\_Internet\\_access](http://en.wikipedia.org/wiki/Satellite_Internet_access)
- [6] TheoryOrg, “Bittorrent Protocol Specification v1.0”, 2008, [Online; accessed 23-November-2008], [Online], Available: <http://wiki.theory.org/BitTorrentSpecification>
- [7] OpenP4P, “What is P4P”, 2008, [Online; accessed 23-November-2008], [Online], Available: <http://www.openp4p.net/>
- [8] SFU, “Modeling and Caching of P2P Traffic”, 2008, [Online; accessed 23-November-2008], [Online], Available: [http://nsl.cs.sfu.ca/wiki/index.php/Modeling\\_and\\_Caching\\_of\\_P2P\\_Traffic](http://nsl.cs.sfu.ca/wiki/index.php/Modeling_and_Caching_of_P2P_Traffic)
- [9] M.Hefeeda, C. Hsu, and K. Mokhtarian. Design and Evaluation of a Proxy Cache for Peer-to-Peer Traffic. School of Computing Science, Simon Fraser University, July 2008.
- [10] PeerApp, “How P2P Caching works”, 2007, [Online; accessed 23-November-2008], [Online], Available: <http://www.peerapp.com/products-ultraband-How-Caching-Works.aspx>
- [11] L. Breslau, P. Cao, L. Fan, G. Phillips, and S. Shenker. Web caching and Zipf-like distributions: Evidence and implications. In Proc. of INFOCOM’99, pages 126–134, New York, NY, Mar. 1999.
- [12] N. Leibowitz, A. Bergman, R. Ben-Shaul, A. Shavit. Are file swapping networks cacheable? Characterizing P2P traffic, Expand Networks, Tel Aviv, 2002.
- [13] M. Hefeeda and O. Saleh. Traffic modeling and proportional partial caching for peer-to-peer systems. IEEE/ACM Transactions on Networking, October 2007. Accepted to appear.
- [14] <http://netcat.sourceforge.net>
- [15] <http://www.stunnel.org>
- [16] <http://www.netfilter.org>

# Mining web logs to improve web site organization and structure

<sup>1</sup>Ján Kliment

<sup>1</sup>Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

<sup>1</sup>te040522@tuke.sk

**Abstract**—This article is about description of knowledge acquiring domains and their division. In later, the principals of web logs mining will be discussed also with descriptions of the problems to be met during mining process. Finally, the oriented graphs and Dijkstra shortest path algorithm as a helpful tool for web logs mining will be discussed

**Keywords**— knowledge, web logs mining, oriented graph, shortest path algorithm.

## I. INTRODUCTION

Today, there is a great need of dynamic and, if possible, automatic development of information systems and technologies. Developers have to use knowledge, techniques of artificial intelligence and AI agents for independent learning and progress of information systems. This systems must be able to develop itself without cooperation with human factor. However, all this is impossible without knowledges. Recently, there are many scientific studies to improve the techniques of knowledge acquiring and processing.

This article represents in short, topics to be discussed in my diploma thesis. Not all of the named experiments are verified in practice, they're still in state of exploration.

The goals of this article are following:

- provide the division of domains for knowledge acquiring
- describe the basics of web logs mining and its utilization
- describe the oriented graphs as a helpful tool for knowledge representation
- apply Djikstra algorithm of shortest way as a tool for acquired knowledge utilization

In short, knowledge can be represented as follows [10]:

Let we have: 18/03/1986

- 1<sup>st</sup> grade –simple string of characters - *data*
- 2<sup>nd</sup> grade – variable (for example *Date*) - *information*
- 3<sup>rd</sup> grade – additional information about it – *knowledge*

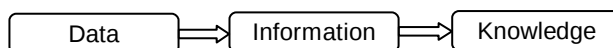


Figure 1: Basic approach of knowledge mining

## II. DOMAINS OF KNOWLEDGE ACQUIRING

There are various knowledge acquiring domains named in the scientific articles. The most inducted criterion for this division is structuring of data to be mined. In common, undistinguished text is being marked as non-structured. A term „*TextMining*“ (acquiring information/knowledge from text) is used in this case. But when we're discussing about structured data, we've to use the term „*DataMining*“ (acquiring information/knowledge from data). *HyperText* (from greek *hyper* and *text*) could be considered as a semi-structured text because of its syntax and semantics [11]. So, Text Mining and HyperText mining are not the same and must be understood as two various mining domains.

There can be several other domains for knowledge mining. In association with previous domains, we should discuss these [1][2][10]:

- acquiring knowledge from text (*Text Mining*)
- acquiring knowledge from databases (*KDD*)
- acquiring knowledge from hypertext (*Hyper Text Mining*)

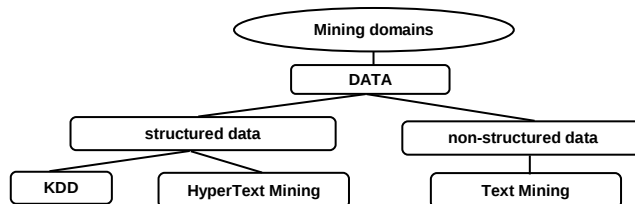


Figure 2: First mining domains division

In common, HyperText mining is divided into three categories [3][11]:

- *Web Content Mining*
- *Web Structure Mining*
- *Web Usage Mining*

The first and second category are oriented to mining knowledge from *hypertext* – the solitary content of web pages (*web content mining*) and set of pages containing additional information situated in hypertext structure (*web structure mining*) [3][10][11].

Third category, Web Usage Mining, is specialized to mine knowledge from another source – *web logs* from the web server [3][10].

According to this, we should now understand *Web Usage Mining* (in next, *Web Logs Mining*) as a distinctive and independent category of knowledge acquiring.

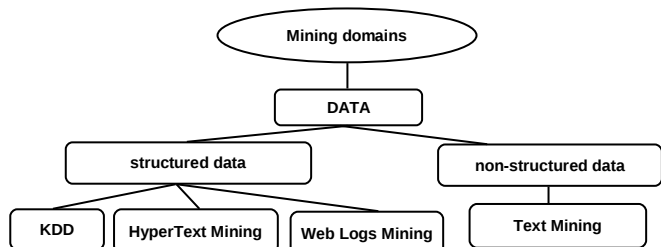


Figure 3: Second mining domains division [10]

### III. MINING WEB LOGS

The goal of web logs mining is to discover *user access patterns* (or *sequences*) and knowledge useful for applying in another domains [2][6].

Web servers generate large volumes of information about usage of web pages. Web logs were originally dedicated for server administrators and security technicians. The most used format of web logs is called *Common Log File Format*. It's structure is following [2][3] [7]:

- requested URL address
- requestor IP address
- timestamp (date and time of request)

The most frequent problems in web logs mining process are [10]:

- using cache on user site (many requests are not sent to web server) [3][6][9]
- using proxy server (problem with recognition of users) [9]
- using backwards operation (problem with recognition of user's web page navigation) [5]
- index and content page problem (problem with page type recognition) [2][5]
- viewing time problem (measurement of user's interests for web pages) [9]

One of the most practical usages of web logs mining should be improving the structure of web pages. Many companies use this for marketing and commerce. But the main goal of web pages should be providing useful information for their visitors. In this case, improving the structure has great importance for the information to be find out effectively [2][3][9][10].

Hierarchical structure of web pages can be a problem due to different opinion of web page creator and its visitors about the page structure. In short, searched document location is not always clear. *Predication* of the next user access is based on mining web logs. This *prediction model* is then used to find out *frequent user access patterns* [4] [6][7].

When user is accessing web pages with some known purposes and intentions, we can expect that this pages contains concrete terms and topics. User then decides

about the relevance of these pages and determines the page quality – *page authority* [2].

The theory of finding out concrete information and its location finally indicates the *model of visitor search patterns* [5]. The basic interpretation of this model is *oriented graph*. Results mined from this graph can be used for creating *adaptive web sites* – the sites automatically improving their structure and inner organization [7].

### IV. ORIENTED GRAPH CONSTRUCTION

Construction of oriented graph representing user's interest on the web is based on following principal [5]:

- searching for single target
  - user's interest for one specific page *T*
- searching for set of targets
  - user's interest for set of pages  $T_1, T_2, \dots, T_n$

The basic algorithm of graph construction [5]:

- start from the root page of web
- if actual location, page *C*, is not the searched target *T*
  - follow the link seems to have the most probability to lead to *T*
- else return to parent page (action with some unknown probability) or give up the finding target page (also with some unknown probability)

The result given by this algorithm is shown on Figure 4.

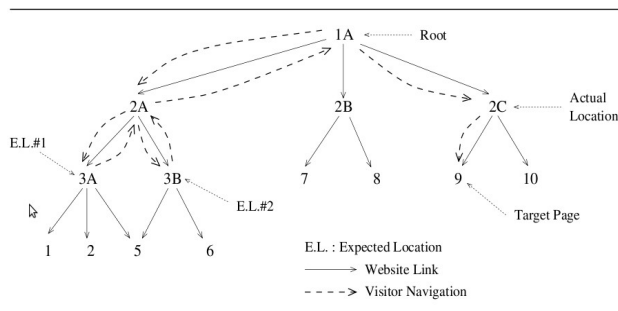


Figure 4 – Oriented graph representing user search patterns and his interests on the web [5]

### V. APPLYING DIJKSTRA ALGORITHM OF THE SHORTEST PATH

Definition and principals of this algorithm, designed by *Edsger Dijkstra* in 1959, are following [8]:

- A graph search algorithm that solves the *single-source shortest path problem* for a graph with nonnegative edge path costs, producing a shortest path tree.
- For a given source vertex (node) in the graph, the algorithm finds the path with lowest cost (i.e. the shortest path) between that vertex and every other vertex. It can also be used for finding costs of shortest paths from a single vertex to a single destination vertex by stopping the algorithm once the shortest path to the destination vertex has been determined.

Oriented graph that is created illustrates all users' page navigation on the web server and shows their interests. We could assume, that leafs of our graph illustrates the pages (or information) which the users were looking for. Of course, it's not always an accurate information – user can visit this page, but if it doesn't belong to the domain of his interest – he simple leaves this page using the backward button. The theory of *viewing time* can be used here – when this time is minimal, we assume, that this page is not interesting for the user and will not be marked as a searched page/information. High count of these *uninteresting* pages with always *minimal viewing time* should also indicate wrong structure and organization of the page – users are looking for something, and they're also expecting the location of it - but actual location is elsewhere.

The paths in oriented graph should be the following:

- *simple path to somewhere*
- *the most used path to concrete page/information*
- *the shortest path acquired by Dijkstra algorithm*

This indicates the real usage and amount of Dijkstra algorithm:

- if all most used paths and shortest paths by Dijkstra algorithm are the same, the page is correct and needs *no improvement of structure and inner organization*
- if many most used paths and shortest paths by Dijkstra algorithm are miscellaneous, the page needs *to be improved – to change the structure according to found shortest paths*

## VI.CONCLUSION

The goal of this article was to illustrate the needs and usage of mining – show the division of these problems and describe one of its domain, Web Logs Mining. There are many studies about this problem so it's impossible to cover the whole scope of it. Only the most known were named in this article.

As we could see there are many problems in the web logs mining and many of them were not resolved till today.

Although this article shows some resolution of the problem of improving web page structure and organization, it's not solving all obstacles of this process – used techniques are not always accurate, and in my opinion, will never be fully automatical – there always will be a need of human factor. It's true that web logs offer many useful information and algorithm of prediction are on very high distinction, but there will always be a need of human decision in some cases.

## VII.ACKNOWLEDGMENT

This work was supported by VEGA Grant No. 1/0350/08 Knowledge-Based Software Life Cycle and Architectures.

## REFERENCES

- [1] Paralič J.: Objavovanie znalostí z databáz, Elfa Košice, 2002.
- [2] Han J., Kamber M.: Data mining - Concepts and techniques, Morgan Kaufmann Publishers, San Francisco, 2001.
- [3] Veččera M.: Dobývání znalostí z webu, Diploma Thesis, FI MU Brno, 2007. (reachable on: <http://is.muni.cz>)
- [4] Chen J., Cook T.: Mining contiguous sequential patterns from web logs, WWW 2007 Conference Bannf & Alberta, Canada, 2007. (reachable on: ACM 978-1-59593-7/07/0005)
- [5] Srikant R., Yang Y.: Mining weblogs to improve website organization, WWW10 Conference, Hong Kong, 2001. (reachable on: ACM 1-58113-348-0/01/0005)
- [6] Yang Q., Zhang H. H., Li T.: Mining web logs prediction models in www caching and prefetching, KDD 01 San Francisco, 2001. (reachable on: ACM 1-58113-391- x /01/08)
- [7] Perkowski M., Etzioni O.: Adaptive Web Sites - Automatically Synthesizing Web Pages. In Proc of the 4th Pacific-Asia Conference on Knowledge Discovery and Data Mining, 2000.
- [8] [http://en.wikipedia.org/wiki/Dijkstra%27s\\_algorithm](http://en.wikipedia.org/wiki/Dijkstra%27s_algorithm)
- [9] Shahabi C., Zarkesh A.M., Adibi J., Shah V.: Knowledge Discovery from Users Web-Page Navigation, In Proc. of the Eight Intl. Workshop on Research Issues in Data Engineering (RIDE), pages 20-29, 1997.
- [10] Kliment J. - Získavanie a spracovanie znalostí, Diploma Thesis Proposal, FEI TU, Košice, 2009.
- [11] Malik F.: Extrakce informací z hypertextu, Diploma Thesis, FI MU Brno, 2007. (reachable on: <http://is.muni.cz>)

# Mobile Wireless Clients Streaming

<sup>1</sup>Pavol KOCAN, <sup>2</sup>Ján MOCHNÁČ

Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

<sup>1</sup>pavol.kocan@tuke.sk, <sup>2</sup>jan.mochnac@tuke.sk

**Abstract**—Video transmitted over wireless channel is always incoherent to packet losses, especially when one or more clients are moving away from access point. One way, how to deal with this problem is estimation of client motion. In this paper we focused on analyzing the client mobility problem with simulations and propose problem solution possibilities.

**Keywords**—multimedia streaming, wireless mobile clients, Opnet.

## I. INTRODUCTION

With growing interest in digital video delivery also grows the interest in providing of sufficient video quality. This regards also Wi-Fi clients where is the specific problem with client mobility that causes delay increase. There are several manners for satisfying quality demands. On the one hand, there can be used error concealment methods which try to approximate losses in video. Another way how to ensure adequate video quality is regulation of the signal power transmitted from access point for the mobile client or its motion prediction.

This paper is divided into 3 main sections. In the beginning we describe impact of mobile clients of wireless networks for streaming quality, next section presents some experimental results and finally short conclusion with some solution recommendation.

## II. WIRELESS NETWORKS AND STREAMING

The main appeal of wireless cellular and ad hoc networks is that they allow both user mobility and untethered connectivity. However, user mobility poses significant challenges to network operations such as routing, resource management, and Quality of Service (QoS) provisioning, especially when it comes to the QoS provisioning of multimedia services [1].

### A. Bandwidth Limitation

Wireless streaming systems are limited by wireless bandwidth and client resources. Client resources are often practically limited by power constraints and by display, communication, and computational capabilities; for example, wireless transmission and even wireless reception alone typically consume large power budgets. In order to make the most efficient use of wireless bandwidth and client resources, it is desirable to send clients the lowest bandwidth video streams that match their display and communication capabilities [2].

Available throughput estimation and rate reduction are ways how to deal with variable bandwidth in wireless transmission. Rate reduction transcoding represents transcoding from a higher bit rate to a lower bit rate and may employ requantization and/or spatio-temporal resolution reduction techniques. Rate reduction transcoding can be further classified into two categories: pixel-domain transcoding and DCT domain transcoding. The performance of pixel-domain transcoding is better than DCT-domain transcoding while the complexity of pixel-domain transcoding is much higher. With regard to wireless video streaming of stored content, there are generally two major issues to consider: the adaptation of the content and the robustness of the transmission [3].

### B. Optimal Transfer Protocol

Wireless channel is characterized by low bandwidth with unpredictable error. Mobile devices, on the other hand, are characterized by their low processing/computational capability and low memory. Traditionally, User Datagram Protocol (UDP) is used for media streaming. However, UDP is not effective for wireless streaming, packets may be lost during transit. To offer good-quality video, these losses have to be mitigated. Retransmission, FEC (Forward Error Correction), and error concealment are techniques which may be used. For video stream, some frames (I frames) and some data fields (synchronization bits) are more important than others and need to be protected. Since wireless error occurs at any time, these important data may be lost, leading to degradation in quality. If those more important frames or data fields can be selectively protected, better video quality would be achieved. TCP is a reliable protocol, and hence effectively addresses the synchronization and retransmission problem. There is no need of complex error concealment and resilience mechanisms which need to be implemented in the client. It is more flexible in choosing which frame to transmit and at what time. No extra framing overhead such as RTP and RTCP is required. It also adapts its transmission rate according to the available network bandwidth, thereof allowing the video applications to make full use of the bandwidth [4].

To improve network connectivity at the network layer, many mobility prediction schemes have been proposed to predict the availability of wireless links in the future, which allows to build more stable end-to-end connections. The provisioning of continuous streaming services of multimedia data in wireless networks, such that streaming interruptions



may be avoided or minimized as much as possible in the user experiences when consuming continuous media is at the application layer [1].

### III. SIMULATIONS

This section describes two simulations at one Opnet scenario. The problem of mobile clients is illustrated in two figures. The network consist of access point (AP) and ten wireless clients. Buffer size at the data transmitter was set on 256 kB, transfer rate 11 Mbps, technique for multiple access was DSSS and the transmission power 1 mW. Both simulations were for the server and 10 wireless devices, four fixed stations and six mobile devices in small LAN network with dimensions 100x100 meters, where three clients received H.263 video stream and the other clients another application

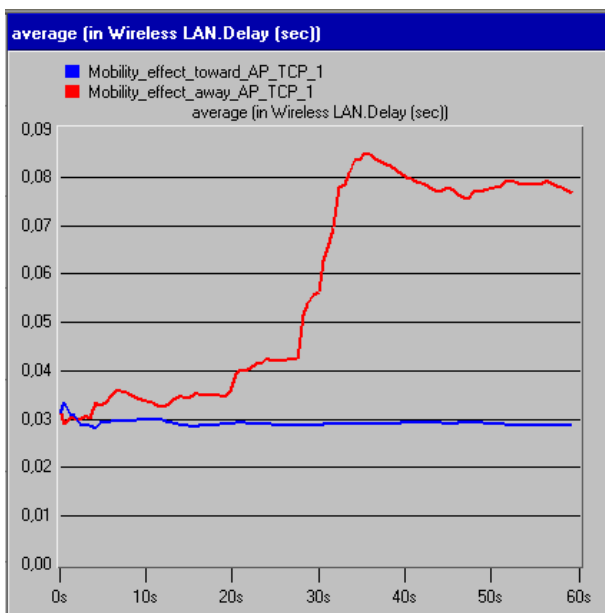


Fig. 1. Delay in seconds for mobile clients moving toward and away from access point

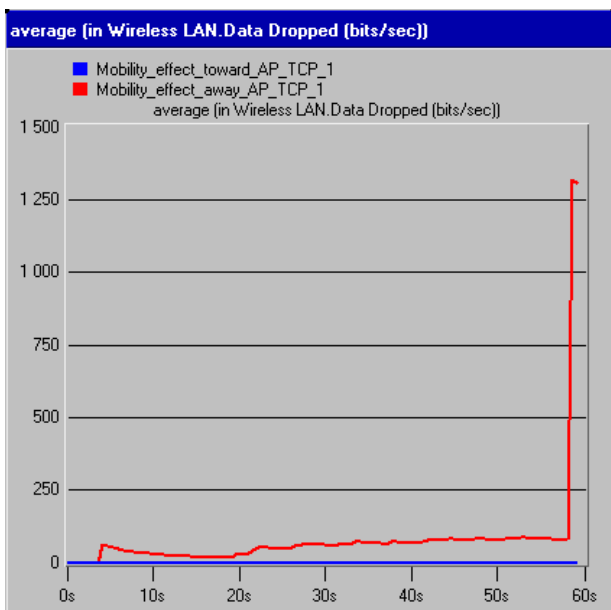


Fig. 2. Data dropped in bits per seconds for mobile clients moving toward and away from access point

like e-mail, file transfer, voice communication over IP and web.

In the Figure 1 delay in seconds for mobile clients moving toward to (blue curved line) and away (red curved line) from access point is illustrated. Delay is approximately around the three hundredth seconds for case of moving toward the access point. As we can see this is value with little change when client is moving toward access point, so the value 0.03 seconds isn't caused by distance of client from the access point. But in case of moving away from access point delay is increasing above 0.08 seconds and stay nearby this value, thereby increase of distance has increases delay value. Different situation is in the Figure 2. While the dropped data for client moving to access point is nearby zero value, but when the distance between access point is increasing then dropped data slowly increase to value of 100 bits per second till the fifty eighth second, when the loss jump above the 1300 bits per second. At this time mobile client reach the network limitation and may cause the visible distortion in video streaming.

### IV. CONCLUSION

In this paper we showed that delay in transfer between access point and clients that moves away from access point may increase very fast. Following dropout after increasing the delay is unpleasant for wireless streaming user and can be solved by throughput estimation and rate reduction but the most efficient could be mobility prediction that is developing and testing for AD-HOC wireless devices in mesh network. The improving methods are available not only at the network layer but also at the application layer.

### REFERENCES

- [1] Baochun Li and Karen H. Wang: NonStop: Continuous multimedia streaming in wireless ad hoc networks with node mobility, IEEE Journal on Selected Areas in Communications, 2003, vol. 21, pp: 1627 -1641
- [2] Susie J. Wee and John G. Apostolopoulos: Secure scalable video streaming for wireless networks; In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, 2001, pp: 2049 - 2052
- [3] Anthony Vetro, Jianfei Cai, Chang Wen Chen: Rate-reduction transcoding design for wireless video streaming, In Proceeding of the International Conference Image Processing, Vol. 1, 2002, pp: 1-29- 1-32
- [4] Chi-fai Wong, Wai-lam Fung, Chi-fai Jack, Tang S. -h, Gary Chan: TCP streaming for low-delay wireless video, Second International Woekshop on Quality of Service in Heterogeneous Wired/Wireless Networks, 2005
- [5] <http://www.opnet.com/solutions/>

# Fuzzy direct torque control of the AM

Štefan KÖVER

<sup>1</sup>Dept. of Electrical, Mechatronic and Industrial Engineering, FEI TU of Košice, Slovak Republic

[kever@inmail.sk](mailto:kever@inmail.sk)

**Abstract**— The paper describes the method of direct controlling of a squirrel cage asynchronous motor torque. Torque control of a squirrel cage AM is realized using Takahashi method of direct torque control that is modified by using fuzzy logic. Fuzzy controller replaces switching table that determines the stator voltage vector. Further, appearing in the paper are waveforms for both methods and in compared in the closing part are simulation results attained by Takahashi method and Takahashi method extended by use of fuzzy logic, respectively.

**Keywords**— Direct torque control (DTC), Takahashi Method, fuzzy logic.

## I. INTRODUCTION

Direct torque control is a relatively young method that was in 1985 publicised by Prof. Manfred Depenbrock for the first time. Hence, the method is named the Depenbrock method, and this became the basis for various modifications such as: Takahashi method, Miro method, or New method. Direct torque control is not utilising PWM (Pulse width modulation) modulator, and thus is the motor response to change in the control signal shortened. [1][4]

The present paper deals with the fuzzy logic adjusted Takahashi method. The switching method used with the original Takahashi method is replaced by fuzzy controller that based on the inputs determines an appropriate vector of the stator voltage. The objective is to attain, by applying fuzzy logic, less overshoot waveforms of torque, and quite naturally to enhance quality of control.

## II. DTC TAKAHASHI METHOD

This direct torque control method is typical by movement of the stator flow vector inside the anuloid, when, in a simplified case, the trajectory resembles a circle. From the voltage equation for stator it holds that [2]:

$$\bar{\psi}_1 = \int (\bar{u}_1 - R_1 \cdot \bar{i}_1) dt \quad (1)$$

Where  $\bar{i}_1$  is vector of the stator current. If we will neglect the stator resistance  $R_1$  than it will hold [2]:

$$\bar{\psi}_1 = \int \bar{u}_1 dt \quad (2)$$

From the above relations it follows that at neglecting the influence of stator resistance the change of the magnetic flux of stator  $\bar{\psi}_1$  is proportional to vector of the stator voltage  $\bar{u}_1$ .

For the torque M it holds:

$$M = \frac{3}{2} \cdot \frac{p \cdot (1 - \sigma)}{\sigma \cdot L_h} \cdot \psi_1 \cdot \psi_2 \cdot \sin \gamma \quad (3)$$

where  $\sigma = 1 - \frac{L_h^2}{L_1 \cdot L_2}$

Where p is number of pole pairs,  $L_1$  inductance of stators,  $L_2$  inductance of rotors,  $L_h$  main inductance,  $\sigma$  is introduce the overall leakage factor,  $\bar{\psi}_2$  is magnetic flux of rotor,  $\gamma$  is angle between magnetic flux of stator and magnetic flux of rotor.

The equation implies that the motor torque depends on mutual position of vector of the stator and rotor flux, respectively.

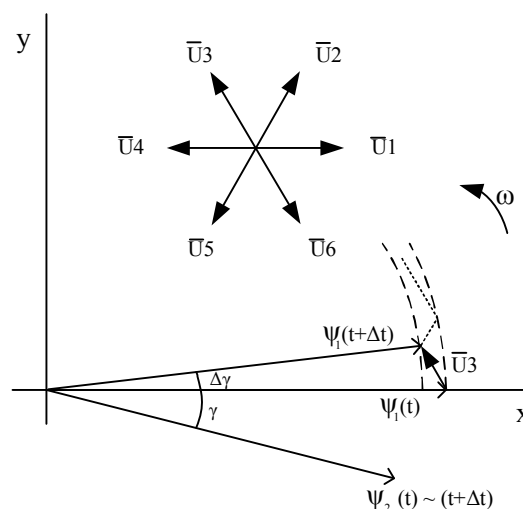


Fig.1 Magnetic flux and torque controlling principle Where  $\omega$  is angular speed.

Principle advantages of the method [3]:

- Prompt torque response,
- Instantaneous values of flux and torque are calculated only from principal quantities,

- Switch is switched so that the torque response would be under transient conditions as fast as possible.

The Takahashi method principle stems in getting an appropriate voltage vector to the motor stator winding. Sensed at the inverter output phases are currents and voltages that are transformed into  $\alpha - \beta$  coordinate system. From these, calculated are  $\alpha$  and  $\beta$  constituents of the stator flux.

So that the stator flux vector moved along the preset trajectory it is necessary to know its magnitude and position along each time interval.

The circle resulting from the stator flux vector's movement is divided into 6 parts, i.e. 6 sectors. Whenever the flux vector moves within a sector, alternately switched are always only two vectors that, moreover, correspond with two neighbouring switching statuses of the inverter.

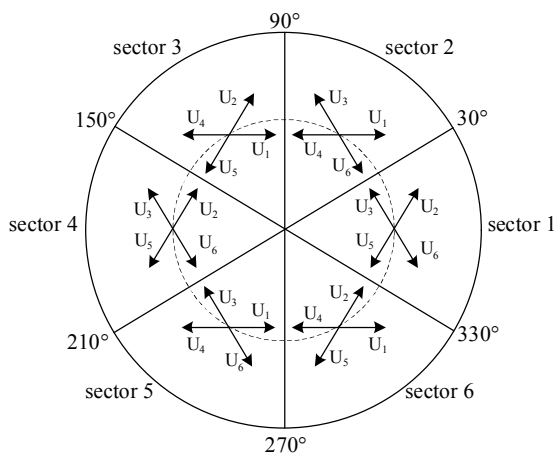


Fig.2 Switching voltages within individual sectors

Control error of the torque and of magnetic flux is brought to hysteretic comparator. The comparator output and the sector within which is the magnetic flux vector located are brought to the address-evaluating block. The block's logic switching circuitry evaluates the address, and this is sent to the switching table. Selected in the switching table is voltage vector the switching combination of which is sent to the inverter that, in turn, connects desired voltage to the stator winding phases.

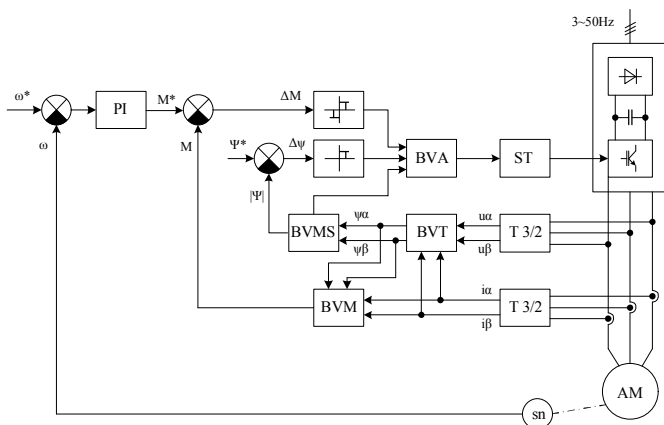


Fig.3 Block diagram of Takahashi direct torque control

PI – PI controller, BVT – flux calculating block, BVA – address evaluating block, BVMS – module of magnetic flux and sector calculating block, T 3/2 – transformation, BVM – torque calculating block, ST – switching table

With this method, still resolved has to be reversal of the motor and its excitation. An option to attain more smoothed waveforms of quantities is to apply fuzzy logic upon selecting the voltage vector.

### III. TAKAHASHI DTC METHOD WITH FUZZY LOGIC

Within this modification of the Takashi direct torque control hysteretic comparators, address evaluating block and the switching table based on which previously determined was the stator voltage vector will be substituted by fuzzy logic.

Based on the inputs the fuzzy controller evaluates the stator voltage vector the switching combination of which will be submitted to the inverter.

The fuzzy controller design breaks up to two stages; defined in the first one are fuzzy sets and their distribution, and established in the second part are rules that assign respective outputs to the inputs.

Block diagram of the method is shown below.

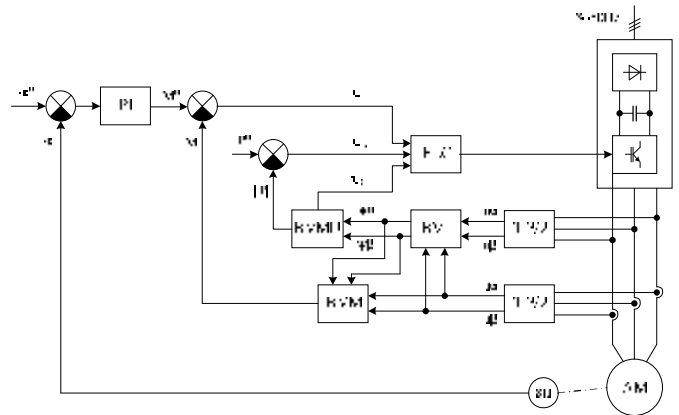


Fig.4 Block diagram of by Takashi method effected fuzzy logic direct torque control

PI – PI controller, BVT – flux calculating block, BVMS – flux and angle module calculating block, T 3/2 – transformation, BVM – torque calculating block, FLC – fuzzy controller

The fuzzy controller is of Mamdani controller type, which is designed and realized within the Fuzzy Logic Toolbox (Matlab) suite. The fuzzy controller inputs are: torque control error  $e_m$ , magnetic flux control error  $e_\psi$ , and the stator flux angle. The FLC output is the stator voltage vector  $u_0$  to  $u_7$ .

The torque control error, magnetic flux control error, and the stator flux angle present linguistic variables in the fuzzy logic. Torque deviation is described by five linguistic values, and error of the magnetic flux is described by three linguistic values. Membership functions are triangularly shaped save of the side ones that attain a trapezoid shape.

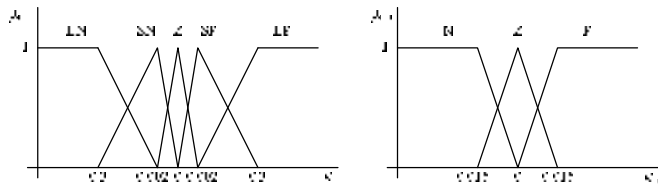


Fig.5 Distribution of the membership functions, of torque and magnetic flux deviations

The sector situated in which is the stator flux vector is evaluated based on its angle.

The angle is described by 7 linguistic values, namely by S1 to S6, whilst S1, i.e. the first vector is divided into two parts – namely into S1.1 and S1.2.

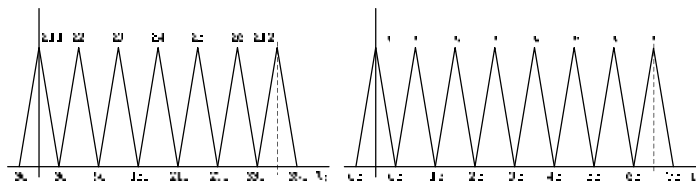


Fig.6 Distribution of membership functions, of stator flux angle and the stator voltage vector

Fuzzy controller operation rests upon fuzzy rules. The advantage of this presentation of knowledge is transparency and ease of human legibility. The rules are of IF-THEN type.

TABLE 1  
TABLE OF FUZZY RULES FOR SECTORS S1-S3

sector	S1.1/S1.2			S2			S3		
	P	Z	N	P	Z	N	P	Z	N
LP	u <sub>1</sub>	u <sub>2</sub>	u <sub>3</sub>	u <sub>2</sub>	u <sub>3</sub>	u <sub>4</sub>	u <sub>3</sub>	u <sub>4</sub>	u <sub>5</sub>
SP	u <sub>2</sub>	u <sub>2</sub>	u <sub>3</sub>	u <sub>3</sub>	u <sub>3</sub>	u <sub>4</sub>	u <sub>4</sub>	u <sub>4</sub>	u <sub>5</sub>
Z	u <sub>1</sub>	u <sub>7</sub>	u <sub>4</sub>	u <sub>2</sub>	u <sub>0</sub>	u <sub>5</sub>	u <sub>3</sub>	u <sub>7</sub>	u <sub>6</sub>
SN	u <sub>6</sub>	u <sub>6</sub>	u <sub>5</sub>	u <sub>1</sub>	u <sub>1</sub>	u <sub>6</sub>	u <sub>2</sub>	u <sub>2</sub>	u <sub>1</sub>
LN	u <sub>1</sub>	u <sub>6</sub>	u <sub>5</sub>	u <sub>2</sub>	u <sub>1</sub>	u <sub>6</sub>	u <sub>3</sub>	u <sub>2</sub>	u <sub>1</sub>

TABLE 2  
TABLE OF FUZZY RULES FOR SECTORS S4-S6

sector	S4			S5			S6		
	P	Z	N	P	Z	N	P	Z	N
LP	u <sub>4</sub>	u <sub>5</sub>	u <sub>6</sub>	u <sub>5</sub>	u <sub>6</sub>	u <sub>1</sub>	u <sub>6</sub>	u <sub>1</sub>	u <sub>2</sub>
SP	u <sub>5</sub>	u <sub>5</sub>	u <sub>6</sub>	u <sub>6</sub>	u <sub>6</sub>	u <sub>1</sub>	u <sub>1</sub>	u <sub>1</sub>	u <sub>2</sub>
Z	u <sub>4</sub>	u <sub>0</sub>	u <sub>1</sub>	u <sub>5</sub>	u <sub>7</sub>	u <sub>2</sub>	u <sub>6</sub>	u <sub>0</sub>	u <sub>3</sub>
SN	u <sub>3</sub>	u <sub>3</sub>	u <sub>2</sub>	u <sub>4</sub>	u <sub>4</sub>	u <sub>3</sub>	u <sub>5</sub>	u <sub>5</sub>	u <sub>4</sub>
LN	u <sub>4</sub>	u <sub>3</sub>	u <sub>2</sub>	u <sub>5</sub>	u <sub>4</sub>	u <sub>3</sub>	u <sub>6</sub>	u <sub>5</sub>	u <sub>4</sub>

IV. COMPARISON OF DTC SIMULATION WAVEFORMS RESULTING FROM USE OF TAKASHI METHOD AND TAKASHI METHOD PLUS FUZZY LOGIC

Simulated motor parameters:

$$L1=0,4862H \quad R1=9,978\Omega \quad J=0,0054kg.m^2$$

$$L2=0,4862H \quad R2=9,933\Omega \quad p=2$$

A. AM WAVEFORMS AT LOW ANGULAR SPEEDS WITH  $Mz=1Nm$

A loaded async motor attained angular speed of 50rad.s<sup>-1</sup>. Next, the angular speed was lowered by 10rad.s<sup>-1</sup>.

Traditional Takashi method was able to follow the desired angular speed till the desired value fell over to the negative. At applying Takashi method with fuzzy logic actual speed

followed the desired value.

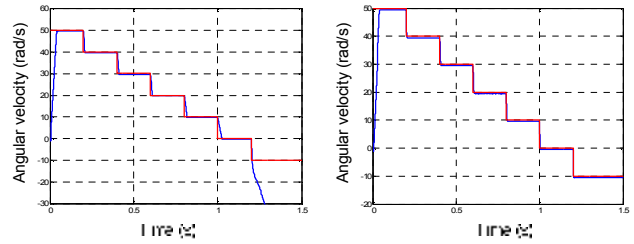


Fig.13 Waveforms of desired and actual angular speed  
a) Takahashi method b) Takahashi method w/ fuzzy logic

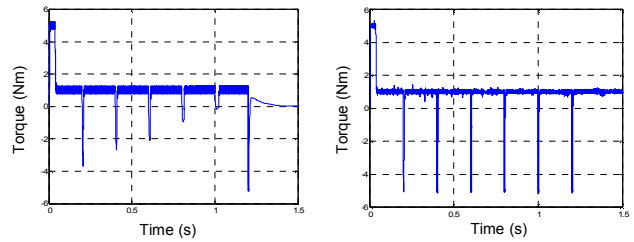


Fig.14 Torque waveforms  
a) Takahashi method b) Takahashi method w/ fuzzy logic

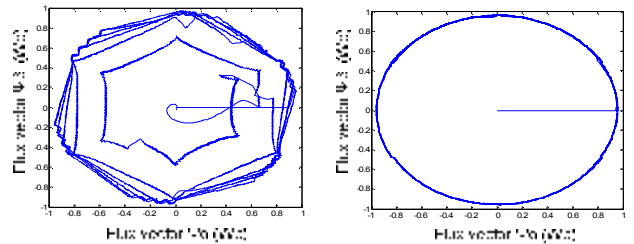


Fig.15 Stator magnetic flux vector waveforms  
a) Takahashi method b) Takahashi method w/ fuzzy logic

V. CONCLUSIONS

The paper deals with direct control of torque of an asynchronous motor using Takashi method extended by fuzzy logic.

With traditional Takashi method the squirrel cage asynchronous motor quantities' waveforms are significantly overshoot. In course of reversal it is necessary to adjust the switching table. At start-up, the motor needs to be excited so that avoided was spiral-wise rise of the stator magnetic flux. The most pronounced the problem is at low angular speeds or at stoppage of a running unloaded motor.

Attained when using Takahashi method with fuzzy logic are less overshoot (approximately 40% lower) squirrel cage asynchronous motor quantities' waveforms. The control does not need any additional treatment during reversal and fuzzy controller excites the motor at start up. Fuzzy logic eliminates also the problem at lower angular speeds or at stopping of a wild running unloaded motor. The angular speed drop is lower when the motor is loaded.

Co-use of fuzzy logic and Takashi method will definitely improve quality of controlling a squirrel cage asynchronous motor.

## ACKNOWLEDGEMENT

The assignment has been solved within the grant VEGA 1/4076/07.

## REFERENCES

- [1] BRANDŠTETER, P.: Střídavé regulační pohony. VŠB TU Ostrava, 1999, ISBN 80-7078-668-X.
- [2] ZEMAN, K. - PEROUTKA, Z. - JANDA, M.: Automatická regulace pohonů s asynchronními motory, Západočeská univerzita v Plzni, 2004, ISBN 80-7043-350-7.
- [3] JOZEFÁK, Lubomír: Priame momentové riadenie s predikčným filtrom, Doktorandská dizertačná práca, Žilinská univerzita, 1998.
- [4] TOUFOUTI, R., MEZIANE, S., BENALLA, H., Direct Torque Control for Induction Motor Using Fuzzy Logic, ACSE Journal, Volume (6), Issue (2), June 2006.
- [5] ŽILKOVÁ, J. - VAŠČÁK, J.: Aplikácie fuzzy logiky a neurónových sietí pri riadení elektrických pohonov: Tempus JEN 02177. Košice : TU, 1996. 126 s.
- [6] PERDUKOVÁ, D. - FEDOR, P.: Fuzzy modelovane v elektrických pohonoch. SYMEP 2006, Plzeň/Nečtiny 13.-15.6.2006. Plzeň : ZČU, 2006. p. 1-9. ISBN 80-7043-455-4.
- [7] VITTEK, J., MAKYŠ, P., ŠTULRAJTER, M., DODDS, S. J., PERRYMAN, R.: Servoposition control with dynamic lag precompensator for PWSM drives, Journal of electrical engineering, vol.7, 2007, Romania, str.: 7.1.4.
- [8] TIMKO, J. - ŽILKOVÁ, J.- GIROVSKÝ, P.: Electrical drives. TU Košice, ISBN 80-8073-529-8, pp 149, Košice 2007.
- [9] BRANDŠTETTER, P.- ŠTEPANEC, L.: Fuzzy Logic Control of Induction Motor Drive.IWCIT'01,VŠB-TU Ostrava,2001,ISBN 80-7078-907-7.
- [10] CIBULA, Ľ.- TIMKO, J. - ŽILKOVÁ, J. – GIROVSKÝ, P.: Direct torque control of the asynchronous motor. In: Journal of Computer Science and Control Systems,(2008),p.18-21.ISSN 1844-6043
- [11] TIMKO, J. – ŽILKOVÁ, J. - GIROVSKÝ, P.: Shaf sensorless vector control of an induction motor. In: Acta Technica ČSAV.Vol.52, no.1 (2007), p.81-91.ISSN 0001-7043
- [12] VITTEK, J. et al.: Comparison of sliding mode and forced dynamics control of electric drive with flexible coupling employing PSMM. In: IEEE International Conference on Industrial Technology IEEE ICIT 2008:21-24 April 2008.Piscataway:IEEE 2008.p.1-6.ISBN 978-1-4244-1706-3.
- [13] ŽILKOVÁ, J. – TIMKO, J. – GIROVSKÝ, P.: Nonlinear system control using neural networks. In: Acta Polytechnica Hungarica.Vol.3,no.4 (2006),p.85-94.ISSN 1785-8860

# Frame Representation of Coherences among UML Model Elements made up of XMI Model Representation

*Ján Kunštár*

Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

jan.kunstar@tuke.sk

**Abstract**— Recent trends in software and system development have revealed the asset of usage of the abstract models through the software life cycle's phases. Great advantage of these models is their simplicity, which predestines them to become the most effective mean for dealing information and knowledge among separate working groups through the software system's life cycle phases. Therefore the abstract models streamline and speed up not only development but suitable models can also improve maintenance process to be more effective and safe. One manner how to utilize software's models for improving the maintenance process is to use them for building knowledge base, which helps to uncover secondary changes, which are the consequence of primary changes required by users of software. Presented paper briefly analyzes the XML Metadata Interchange representation of UML model, which is the source for creation of proposed frame representation. Proposed frame representation presents one part of mentioned knowledge base, which is used for identification of coherences among UML model elements. Proper identification of these coherences is crucial for uncovering secondary changes at the base of performed primary changes. With the entire knowledge base the maintainer is able exposes chain reaction of secondary changes which are consequence of required, primary changes further before required changes are really performed to software system's source code. By this manner the costs needed for proper system's change is markedly reduced.

**Keywords**— knowledge, software maintenance, UML, XML Metadata Interchange (XMI).

## I. INTRODUCTION

In the present, software maintenance task is the most expensive part of the software system life cycle (the surveys indicate that it consumes 60% to 90% of the total life cycle costs [1], [2], [3]). Program comprehension, impact analysis and regression testing are the most challenging problems of software maintenance [2].

An inconsistent state of the software artifacts markedly contributes to all three mentioned problems. When a software system is changed, each artifact which is influenced by this change has to be modified for preservation of software maintainability. Otherwise software artifacts are in inconsistent state – they do not describe the same system.

Knowing and understanding of software system and uncovering the dependences among system's components are the crucial phases during maintenance process, because if

maintainer doesn't know the system, he can't change it properly. One of the core problems over system change are unwanted side effects of performed changes, which where required.

For better understanding of chain reaction lunched by certain required change, were defined two types of changes:

**Primary change** – the change of the system required at the basis of stakeholder objectives or change of system's environment.

**Secondary change** – the change of the system required for elimination of the side effects bring out by primary or another secondary change (changes).

During the entire process of uncovering all required secondary changes, secondary change uncovered in one taxonomy level becomes change which requires other secondary changes. Therefore was defined one more change type:

**Initiatory change** – it is the primary or secondary change, whose realization requires other secondary changes.

Knowledge about coherences among system's components is required for uncovering the unwanted effects of primary changes and subsequent identifying required secondary changes for elimination of these effects.

In this paper the proposed frame representation of software system's UML model is presented. Frame representation, used for identification of coherences among UML model elements (and also among software's components, which are represented by UML elements), is build at the base of XMI representation of UML model.

Proposed frame representation as a part of an entire knowledge base, used for streamlining maintenance process is constituent of Model-Driven Maintenance (MDM) concept, one useful aspect of knowledge-based software life cycle oriented to better usability of all analysis, design and implementation models in maintenance of systems [4].

## II. XML METADATA INTERCHANGE

### A. XMI Essentials

The XML Metadata Interchange (XMI) is a standard for exchanging metadata information via Extensible Markup

Language (XML) specified by Object Management Group<sup>1</sup> (OMG).

It can be used for any metadata whose metamodel can be expressed in Meta-Object Facility (MOF), another standard specified by OMG [5].

One purpose of XMI is to enable easy interchange of metadata between UML-based modeling tools in distributed heterogeneous environments [6]. XMI is in the present the standard for UML model interchange, which is accepted by all significant vendors of CASE tools, which are capable to generate XMI representation of particular model.

The facts, that XMI is the information technology standard, and that it could be generated by majority of widely used CASE tools, are the reasons why the XMI format was selected as the source for proposed frame representation.

Advantages of XMI are [6], [7], [8]:

- XMI is now an international standard: ISO/IEC 19503:2005 Information technology
- The main purpose of XMI is UML model interchange, where UML is standard for system modeling, so we have one common source of information for interchange
- UML metamodel is highly defined specification, managed by OMG, which is nearly generally accepted as independent part of specification
- UML metamodel could be extended and XMI is able to include these extends

### B. XMI Production

XMI's XML document production process is defined as a set of production rules. When these rules are applied to a model or model fragment, the result is an XML document. The inverse of these rules can be applied to an XML document to reconstruct the model or model fragment. In both cases, the rules are implicitly applied in the context of the specific metamodel for the metadata being interchanged [7].

The XMI production rules are defined for elements of three MOF 2.0 packages: Abstractions, EMOF (Essential MOF) a CMOF (Complete MOF), all could be found in MOF to XMI mapping specification [7]. These rules are defined through the mapping of model elements to serialization patterns of serialization model (Fig. 1).

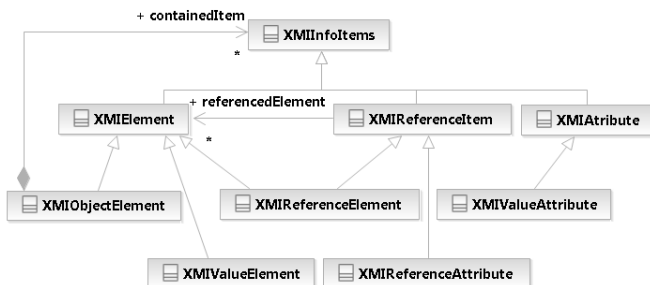


Fig. 1. Serialization model.

An **XMIObjectElement** is an XML element that can contain other information items (XML elements and attributes). An **XMIValueElement** is an XML element that can have a value, but cannot contain other XML elements or attributes. An

**XMIReferenceElement** is an XML element with an *idref* or *href* attribute that references another **XMIElement**. An **XMIReferenceAttribute** is an XML attribute that references an **XMIElement** by *id*. An **XMIValueAttribute** is simply an XML attribute with a value [7].

## III. FRAME REPRESENTATION

### A. Definition of Frame Representation

*Definition 1:*

Let  $n(e)$  be a name of element  $e$  in XMI representation:  $n(e) = QName(e)$ ;  $e \in XMIElement$ , next let  $id(e)$  be an identifier of element  $e$  in XMI representation:  $id(e) = value(id)$ ;  $e \in XMIElement$ , let  $N_A(e)$  be a set of permissible names of attributes of element  $e$  in XMI:  $N_A(e) = \{QName; QName \in (XMIValueAttribute(e) \cup XMIReferenceAttribute(e))\}$ , and let  $V_A(e)$  be the set of permissible values of attributes of element  $e$  in XMI:  $V_A(e) = \{value; value \in (XMIValueAttribute(e) \cup XMIReferenceAttribute(e))\}$ , then ordered trinity  $f(e) = (n(e), id(e), S(e))$  is the frame representing element  $e$ , where  $n(e)$  is name of the frame,  $id(e)$  is frame's identifier,  $S(e)$  is the set of frame's slots except slot *id*:  $S(e) = \{(s_N, s_V); s_N \in N_A(e), s_V \in V_A(e)\} - (id, id(e))$ .

*Definition 2:*

Let  $E = \{e; e \in XMIElement\}$  be the set of all elements in XMI representation of UML model, then the set of ordered trinities  $F = \{(n(e), id(e), S(e)); e \in E\}$ , is complete set of the frames of UML model.

*Definition 3:*

The frame representation of UML model, containing the set of elements  $E = \{e; e \in XMIElement\}$  in XMI, is the ordered pair  $FR = (F, C)$ , where  $F = \{(n(e), id(e), S(e)); e \in E\}$  is complete set of frames of UML model, the set  $C$  is set of coherences among frames  $C \subset \{(t, f_S, f_E); f_S \in F, f_E \in F, f_S \neq f_E\}$ , and  $f_S$  is starting frame,  $f_E$  is ending frame of certain coherence and  $t$  specifies coherence's type.

*Definition 4:*

Let  $FR = (F, C)$  be the frame representation of UML model and suppose existence of coherence  $c \in C$  such, that starting frame of this coherence represents the UML element, which is constituent of another UML element, represented by ending frame, then this coherence is coherence of ISPO (is part of) type.

*Definition 5:*

Let  $FR = (F, C)$  be the frame representation of UML model and suppose existence of coherence  $c \in C$  such, that starting frame of this coherence represents the UML element, whose identifier is value of reference attribute of another UML element, represented by ending frame, then this coherence is coherence of IRPO (is referenced part of) type.

<sup>1</sup> OMG™ is an international, open membership, not-for-profit computer industry consortium, <http://www.omg.org/>.

### B. Frame Representation Model

The next figure (Fig. 2) presents the model of proposed frame representation.

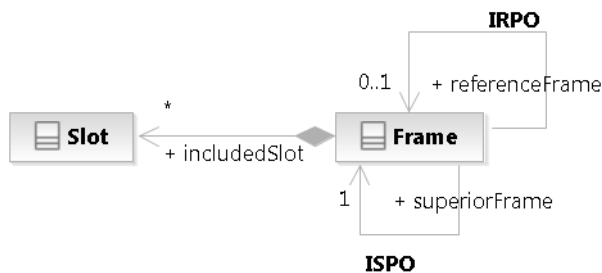


Fig. 2. Model of frame representation.

### C. Mapping from XMI to Frame Representation

Rules for production of frame representation of UML model are defined for serialization patterns of serialization model (Fig. 1). In next table (Table 1) are introduced all rules required for production of frame representation.

TABLE I  
RULES FOR FRAME REPRESENTATION

Instance of XMI element	Frame representation
$e \in \text{XMIObjectElement}$	If the element $e$ is constituent of another XMI element, create: <ol style="list-style-type: none"> <li>1. Frame, which represents element <math>e</math> - frame's values, the same as values of its slots are assigned according definition (Def.1)</li> <li>2. Coherence <i>ISPO</i>, where the starting frame is actually created frame, which represents element <math>e</math> and the ending frame is the frame, which represents superior element to element <math>e</math> in XMI</li> </ol> else create: <ol style="list-style-type: none"> <li>1. Frame, which represents element <math>e</math> - frame's values, the same as values of its slots are assigned according definition (Def. 1)</li> </ol>
$e \in \text{XMIValueElement}$	Create: <ol style="list-style-type: none"> <li>1. Frame, which represents element <math>e</math> - frame's name assign according definition (Def. 1) - generate unique frame's identifier - <math>S = \{(value, value(e))\}</math></li> <li>2. Coherence <i>ISPO</i>, where the starting frame is actually created frame, which represents element <math>e</math> and the ending frame is the frame, which represents superior element to element <math>e</math> in XMI</li> </ol>
$e \in \text{XMIReferenceElement}$	If the element $e$ is constituent of another XMI element, create: <ol style="list-style-type: none"> <li>1. Frame, which represents element <math>e</math> - frame's values, the same as values of its slots are assigned according definition (Def. 1)</li> <li>2. Coherence <i>ISPO</i>, where the starting frame is actually created frame, which represents element <math>e</math> and the ending frame is the frame, which represents superior element to element <math>e</math> in XMI</li> <li>3. Coherence <i>IRPO</i>, where the starting frame is the frame, which</li> </ol>

	is the target of reference through the reference attribute from actually created frame, which represents element $e$ in XMI. The ending frame is actually created frame, which represents element $e$ in XMI. else create: <ol style="list-style-type: none"> <li>1. Frame, which represents element <math>e</math> - frame's values, the same as values of its slots are assigned according definition (Def. 1)</li> <li>2. Coherence <i>IRPO</i>, where the starting frame is the frame, which is the target of the reference through reference attribute from actually created frame, which represents element <math>e</math> in XMI. Ending frame is actually created frame, which represents element <math>e</math> in XMI.</li> </ol>
$a \in \text{XMIValueAttribute}$	Is represented as slot ( $s = (QName(a), value(a))$ ) of appropriate frame, which is closely specified by the value of this slot.
$a \in \text{XMIReferenceAttribute}$	Is represented as slot ( $s = (QName(a), reflD(a)) \vee s = (QName(a), URIref(a))$ ) of appropriate frame, which is closely specified by the value of this slot.

#### Remark:

At the creation of the frame representing XMI element of type *XMIValueElement* it is required to generate an unique identifier for the frame in the frame representation, since this element doesn't contain *id* attribute.

On the next figure (Fig. 3) three different representations of the same class of UML 2.0 model are presented. Figure 3a presents class *Person*, which is the constituent of UML model created in CASE tool (IBM Rational Software Architect 7.0). Figure 3b shows XMI representation of class *Person*, generated by the same tool. The last part of this figure (3c) presents frame representation of UML model, which contains class *Person*.

### IV. CONCLUSION

In this paper, the frame representation of UML model was presented. Proposed frame representation is important part of entire knowledge base used for uncovering secondary changes, which are required as a consequence of performed initiatory changes during the maintenance phase of software systems' life cycle.

The main task of frame representation in the process of uncovering all secondary changes is to identify coherences among frames, which represent elements of UML model of particular software system. Proposed representation is apposite for mentioned task, because it permits to identify taxonomy of UML model's elements; in spite of the fact that this representation retains all knowledge about model's elements, which could be acquired from XMI, it is aimed to identification of coherences among elements; and to top it all it has elementary structure, which eases its implementation.



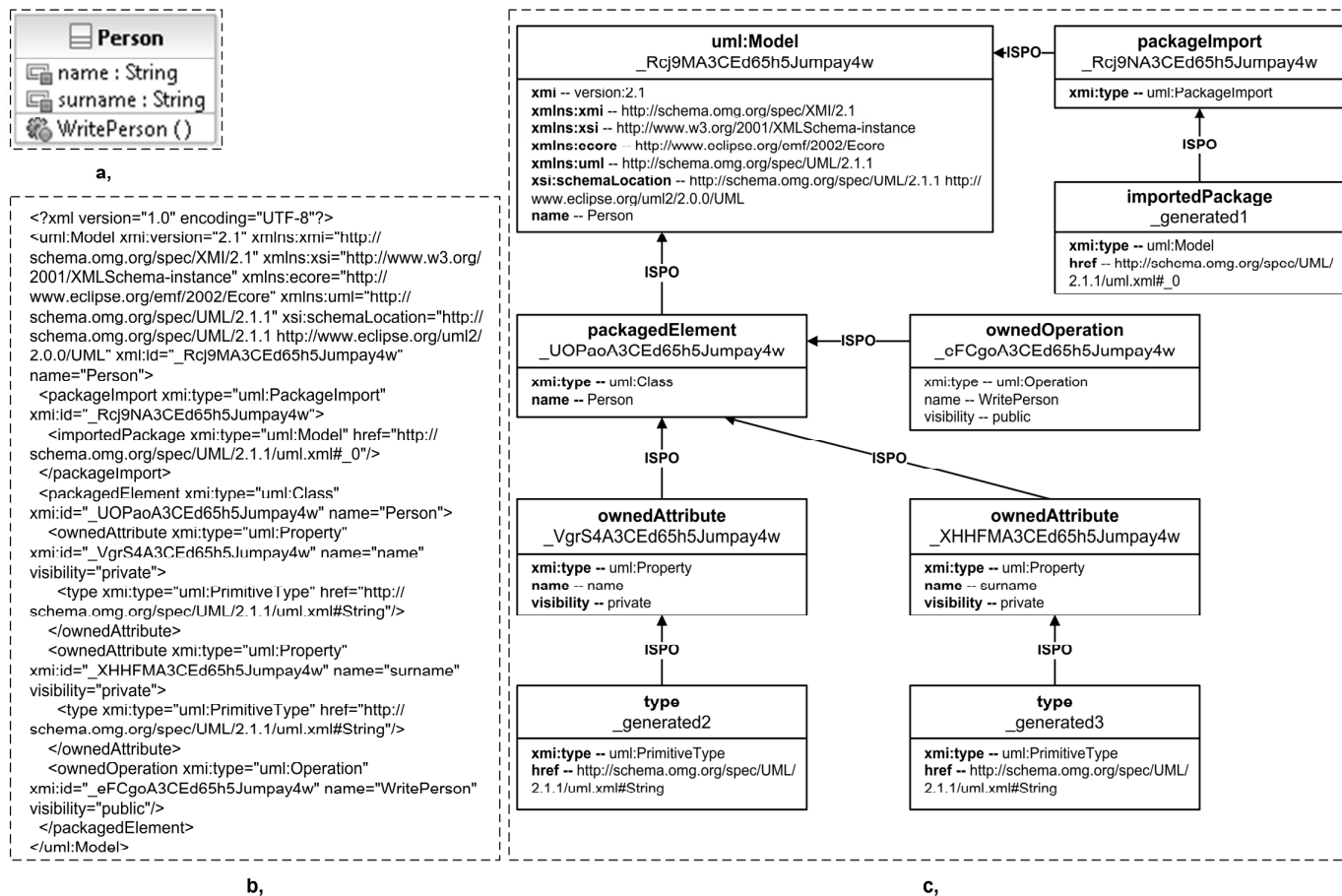


Fig. 3. Three different representations of one model: 3a – class diagram created in CASE tool RSA 7.0, 3b – XMI representation of UML model, 3c – frame representation of UML model.

ACKNOWLEDGMENT

This work was supported by VEGA Grant No. 1/0350/08 Knowledge-Based Software Life Cycle and Architectures.

REFERENCES

[1] J. Kunštár, I. Adamuščinová, Z. Havlice, *The use of development models for improvement of software maintenance*, Acta Universitatis Sapientiae, Informatica, 1, 1, 2009, pp. 45-52, ISSN 1844-6086.  
 [2] K G. Canfora, A. Cimitile, *Software Maintenance. Handbook of Software Engineering and Knowledge Engineering*, volume 1. World Scientific, 2001, ISBN: 981-02-4973-X.

[3] A. P. Grubb, A. A. Takang: *Software Maintenance: Concepts and practice*, SE, ISBN 978-981-238-425-6, World Scientific 2003.  
 [4] J. Kunštár, I. Adamuščinová, Z. Havlice, *Model-Driven Life Cycle, CSE'2008 International Scientific Conference on Computer Science and Engineering*, High Tatras - Stará Lesná, Slovakia, September 24-26, 2008, pp. 255-262, ISBN 978-80-8086-092-9.  
 [5] *Meta Object Facility (MOF) Core Specification*, OMG Available Specification, Version 2.0, formal/06-01-01, January 2006.  
 [6] M. Alanen, I. Porres, *Model Interchange Using OMG Standards*, Proceedings of the 2005 31st EUROMICRO Conference on Software Engineering and Advanced Applications (EUROMICRO-SEAA'05), 0-7695-2431-1/05, 2005.  
 [7] *MOF 2.0/XMI Mapping*, Version 2.1.1, OMG Specification, ptc/07-10-06, December, 2007.  
 [8] *UML 2.0 Diagram Interchange Specification*, OMG Adopted Specification, ptc/03-09-01, September 8, 2003.

# Computational Intelligence in Font Design

*Miron KUZMA*

Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

miron.kuzma@tuke.sk

**Abstract**—This paper gives an introduction to Computational Intelligence application in the domain of font design and basic introduction to Interactive Evolutionary Computation. We test its abilities and practical usability. We look at the domain of font design and describe some experiments with font design. We briefly describe our designed system and analyze its performance. We conclude that proposed system is a possible direction for future font design applications.

**Keywords**— font design, interactive evolutionary computation

## I. INTRODUCTION

Nowadays there are few approaches using Computational Intelligence (CI) in the font design domain. Ian Butterfield and Matthew Lewis [6] worked on a parametric font definition. The basics of their approach are the letters that are individually deformed by collections of implicit surface primitives in 3D modelling software called Houdini. The parametric representation is used in the interactive evolutionary design system to breed fonts. Another approach was performed by Dolinský and Takagi [3]. This approach uses handwritten characters to synthesize fonts and neural network learning. Another application - Fontifier [4][5] - also uses handwriting to create one's personal font.

There are three different approaches in the font design field. The first is the evolutionary approach, the second approach uses neural network and the third one converts users handwriting directly to the digital form.

The reason to synthesize handwriting and to create fonts is that the synthesized handwritten characters enable us to personalize font for the user. It shows one's personal handwriting style. The user is able to set the style of various documents according to his personal font, e.g. writing an e-mail, by using chat program, writing a blog and by other activities. From the recipient point of view, when he gets an e-mail or message written using personal font, he might feel closer to the sender and vice versa. We can say handwriting or "personal font" adds a feeling of personal touch.[3]

## II. INTRODUCTION TO INTERACTIVE EVOLUTIONARY COMPUTATION

One of the future directions of computational intelligence is humanized computational intelligence. One of such technologies is Interactive Evolutionary Computation (IEC). The term we explain in the following part of the paper. As we will see this research domain is famous

with many of its successful applications, the field of its potential application is wide.

The article published by Takagi in 2002 [1] gives a survey of the Interactive Evolutionary Computation (IEC). There exists a large variety of systems using IEC, eg. [7],[8],[9] in image processing, and other system [9] in media database retrieval.

IEC is commonly used in artistic field, engineering field, and other fields. The research categories are: graphic art and computer generated animation, 3D computer generated lightning desing, music, editorila design, industrial design, face image generation, speech processing, hearing aids fitting, virtual reality, database retrieval, data mining, image processing, control and robotics, internet, food industry, geophysics, art education, writing education, games and therapy, social system. Another topic is the research of user interface. It focuses on human fatigue and tries to reduce its unwanted impact.

Interactive Evolutionary Computation is a technique that involves evolutionary computation consisting of genetic algorithms (GA), evolutionary strategy (ES), evolutionary programming (EP), and genetic programming (GP). It aims to optimize the target system based on human subjective evaluation. Regular optimization methods can be used if the specifications or design goal of the target system is numerically given. However, there are many cases that the system performance is not measurable and only human can evaluate the system performance, for example, maximizing sound quality of a hearing aid for a user, generating computer graphics for my living room, generating Jazz-like music. Subjective evaluation includes both KANSEI scale such as preference and evaluation based on domain knowledge.[1]

The Interactive Evolutionary Computation as an optimization method involves Evolutionary Computation (EC). It is a method that uses subjective human evaluation. It is an EC technique thats fitness function is replaced by a human user, because we cannot provide the system with the deterministic/distinct function.

Figure 1 shows a general IEC system where the system output is shown to the user and user evaluates system outputs. The EC optimizes the target system to obtain the preferred output based on the user's subjective evaluation. The IEC technology embeds in the target system following: human preference, intuition, emotion, psychological aspects. We call these using a more general term KANSEI.

There is another important aspect in IEC. It is active user intervention (AUI) which shall motivate the users and shall speed up EC convergence.

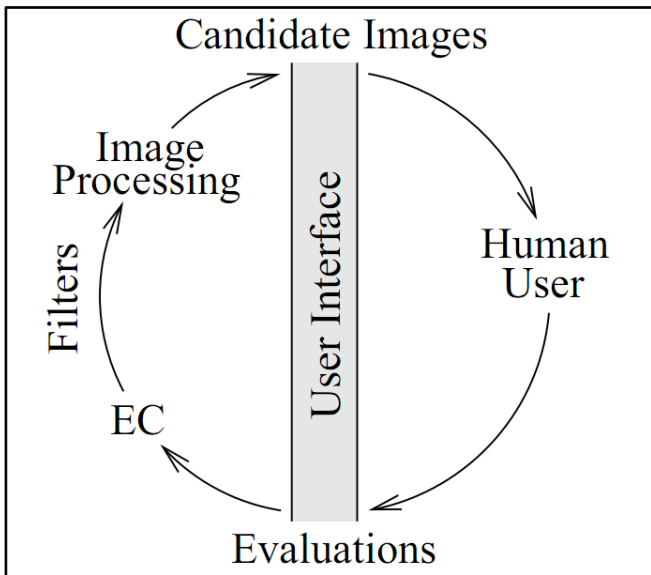


Fig. 1. General IEC system: system optimization based on subjective evaluation[7].

The convergence speed up finally results by direct reduction of human fatigue. We explain the AUI function on the face reconstruction system. When user perceives that a certain facial feature of an individual montage image will improve an EC search, the genes (parameters) corresponding to this facial feature are masked to prevent its change. This masking means that the dimensional number of the searching space is reduced and we are exploring the reduced dimension searching space. That way we are able to accelerate the IEC search.[2]

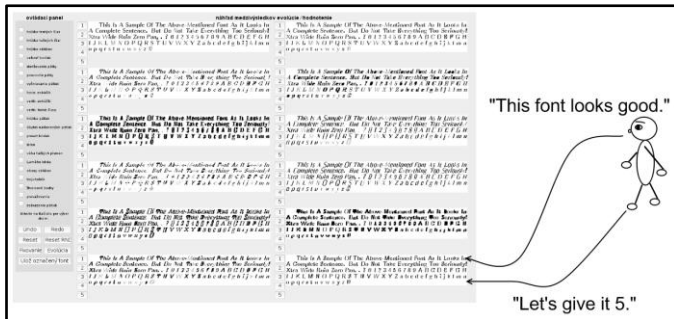


Fig. 2. Font Evolving System: User Interface. The "CONTROL PANEL" with actions on the left side and "SAMPLES PANEL" with Font Samples and their marking buttons on the right side. Evaluating samples on the right side and selecting the action from the control panel is the usual order of steps by IEC programs.

### III. EXAMPLE SYSTEM

Our approach focuses on designing and implementing system that is able to help user to create a font, reduce the time needed for this process or give a basic idea of font to start with. It is using some methods widely spread in evolutionary computation, namely interaction and active user intervention.

The user interface is based on any internet browser, the application was designed as a web application. The Idea of the system complies to IEC basics. On Figure 2 we have a user interface that is split into two main parts: the "CONTROL PANEL" and the "SAMPLES PANEL". The control panel is on the left side and gives us control over the designed system – The Font Evolving System (FES). We have different available actions to change the population – the

fonts. The samples panel on the right side has 12 Font samples with buttons having marks from mark 1 - the worst - to the mark 5 - the best. Those marks have influence on the global evolution process. User has to evaluate the Font samples. He has to click on the mark that corresponds to his own preferences and intentions to evaluate the samples. The next step of the evolution process is to consider the preferable action and click on the corresponding button.

This system focuses also on independence of modules and - what is new - it counts with the future multi-user environment, such as to store the user information, settings and results in database. The lifecycle of the FES is on Figure 3.

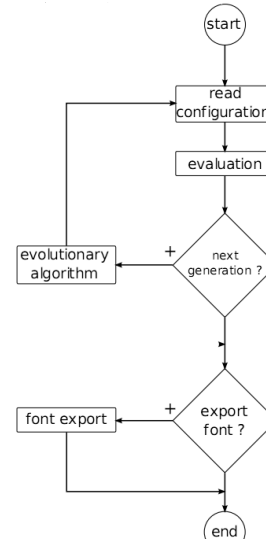


Fig. 3. FES lifecycle.

The modular structure of the designed system was: the module for the user interface, the module for genetics, the module for miscellaneous utility functions, the module for configuration for every user and the global communication module that handled all the other modules.

We took the Computer Modern font as the basis for the system. The font has 62 parameters, we chose 25 from them and experimentally estimated the range for every parameter. So the final space search was a 25-Dimensional space.

We ran experiments with the system to justify its usability among users. Experiments were compared to the manual font design taking into account time, user-friendliness, result – the designed font.

### IV. SUBJECTS TEST ON THE PROPOSED FONT DESIGN SYSTEM

The experiment was done under following conditions. We took 10 subjects and they had been given two tasks. Both task were to design a font which they like. The first was to design a font with the Font Evolving System, and the second task was to manually design a font. However the manual creation of font involves some metafont knowledge. That is why we designed a special interface. The only activity in the second task was to change the values of the parameters in text boxes and confirm them with the "CREATE FONT" button in order to achieve the result. The resulting font is on the right side of the users screen. The first task ran with the designed user interface of The Font Evolving System with its corresponding controls described in the modules

section of the Font Evolving System. The subjects had been given a short user’s manual of the Font Evolving System. For both tasks users had been given two URLs, one for the Font Evolving System and one for the manual creation. That way they were able to run the experiment whenever and wherever they wanted in order to give them maximum of comfort. The initial sample was the same for all of the subjects. As a reward for participating on the experiment the users had the ability to export the designed font to Postscript Type1 format - \*.pfa file.

Finally the subjects had to answer with ”YES” or ”NO” a simple questionnaire containing following 9 questions:

1. Is the semi-automatic font designing better than manual designing of font?
2. Are you satisfied with the result of the semi-automatic font design?
3. Are you satisfied with the result of the manual font design?
4. Do you think you might use the designed font?
5. Do you think others might use the designed font?
6. Are you satisfied with the amount of time spent by the semi-automatic font design?
7. Are you comfortable with the program control?
8. Are you comfortable with the abilities of the program?
9. Are you comfortable with the description of all the available actions in the program?

We obtained 10 filled-out questionnaires. The questionnaire summary is in Figure 4. From the questionnaire structure, the questions 1 to 5 are about the characteristics of the designed algorithm.

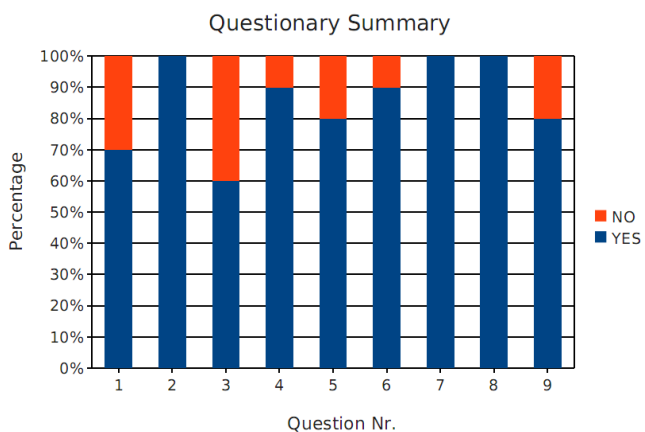


Fig. 4. Questionnaire Summary.

The subjects designed very variable fonts. This was expected, as the Task #1 was to design a font which they like. This is very subjective for every participant as different people have different concepts in the art domain. The general feeling from the users answers in the questionnaire is that such a software can meet the need of the majority of the users that need a software for a font design. The Commercial application of FES will probably require more actions, functions and features, but this should be no obstacle according to our experience with our FES implementation.

A little more than half of the users (70% of the subjects acc. to question #1) found the semi-automatic design better than the manual font design. However all the participants found

the results of FES acceptable, but only 60% of subjects were satisfied with their manually designed results.

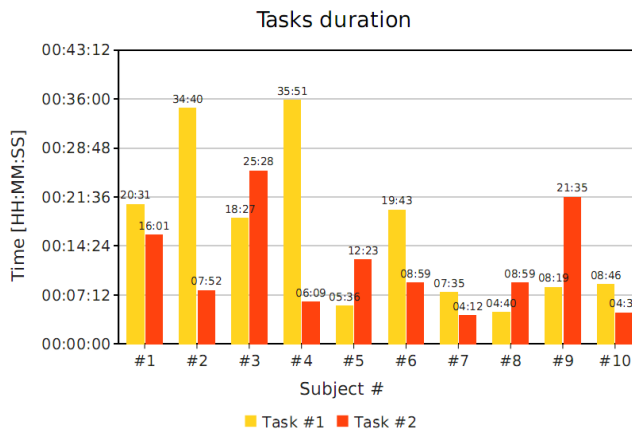


Fig. 5. Tasks Duration.

The time they spend to create a font varies depending of the user’s desires.

We think the presence of the Active User Interaction in the FES is a good idea. Its presence is very helpful - it shortens the time to get an acceptable result. Although we did not include a question about the AUI in the questionnaire, many of subjects were satisfied with the AUI incorporation to our system. The AUI presence in this application has has another good purpose: it enables the user to actively interact with the evolution process and that way enforce the direction of evolution.

V. CONCLUSION

We conclude that proposed system is a possible direction for future font design applications. Our evaluation with subjects showed applicability of this approach.

We surveyed another three approaches in CI with application in the font design domain. The neural network approach by [3], the evolutionary approach by [6] and deterministic approach by [4],[5]. Our system is an approach that belongs to the interactive evolutionary computation.

For the future work we recommend to further study the active user intervention interface and implementation and customization if displayed font samples to improve the comfort of the user.

REFERENCES

- [1] TAKAGI, Hideyuki 2002. Interactive Evolutionary Computing: Fusion of the Capacities of EC Optimization and Human Evaluation In: Proc. of 7th Workshop on Evaluation of Heart and Mind, KitaKyushu, Fukuoka, (November 8-9, 2002)(in Japanese) pp.37-58
- [2] TAKAGI, Hideyuki 2000. Active User Intervention in EC Search In: International Conference on Information Sciences (JCIS2000,) Atlantic City, NJ, USA, (Feb. 27 - Mar. 3, 2000) pp.995-998
- [3] DOLINSKÝ, Ján; TAKAGI, Hideyuki 2007. Synthesizing Handwritten Characters Using Naturalness Learning In: 5th International Conference on Computational Cybernetics (ICCC2007), Gammarth, Tunisia, (Oct. 19-21, 2007) pp.101-106
- [4] Scrapbooking 2008.Turning Your Handwriting Into a Font for Scrapbooking In: <http://scrapbooking.about.com/library/weekly/aa062004a.htm>, achievable on-line. (May 8, 2008)
- [5] Scrapbooking 2008.Turning Your Handwriting Into a Font for Scrapbooking In: <http://scrapbooking.about.com/library/weekly/aa062004a.htm>, achievable on-line. (May 8, 2008)

- [6] Fontifier 2008. Your Own Handwriting on Your Computer! In: <http://www.fontifier.com/>, achievable online. (May 8, 2008) [6]  
FontForge 2008. FontForge: An Outline Font Editor under GPL license In: <http://edt1023.sayya.org/fontforge/overview.html>, <http://edt1023.sayya.org/fontforge/generate.html>, achievable online. (May 8, 2008)
- [7] LEWIS,Matthew2008.EvolvingFonts In: <http://accad.osu.edu/~mlewis/AED/Fonts/>, <http://accad.osu.edu/research/scientificvisualization/htmls/metavolve.htm>, achievable online. (May 8, 2008)
- [8] JAKŠA, Rudolf; TAKAGI Hideyuki 2003. Tuning of Image Parameters by Interactive Evolutionary Computation In: Proc. of 2003 IEEE International Conference on Systems, Man & Cybernetics (SMC2003), Washington D.C., (October 5-8, 2003) pp.492-497
- [9] JAKŠA, Rudolf; TAKAGI Hideyuki; NAKANO Shota 2003. Image Filter Design with Interactive Evolutionary Computation In: Proc. of the IEEE International Conference on Computational Cybernetics (ICCC2003), ISBN 963 7154 175, Siofok, Hungary (August 29-31, 2003)

# Knowledge about Software Architecture

M. Lakatoš\*

\*Technical University of Košice, Department of Computers and Informatics, Košice, Slovakia

lakatos.matej@hotmail.com

**Abstract**— During more and more growing up of complexity of software systems also raise sights on easy handle and understandability of whole system. The easiest way how to reach these destinations from view of programmers is to try to get wrote code-source on higher level of understandability and to connect them with an analyst view which usually use various kind of modeling tools based of UML technology. A programmer and an analyst usually have different knowledge about how the final software should to work and in the event of additional requests on the software system also how and which all parts of the system have to be changed. This article analyzes present state most used technologies during the software development from view of a programmer and an analyst and possible relations between them to keep knowledge about software system together with understandable state and be documented. Advantages and disadvantages of technologies used for software design are analyzed in the paper.

**Keywords**—knowledge, software system life cycle, information systems architecture, abstract model, knowledge-based techniques

## I. INTRODUCTION

During the last couple decades, complexity of software systems grown up staggeringly. Pfleeger [1] in 1998 did survey on more then 8000 applications and discovered that common reasons of break down software projects are :

- uncompleted or no understandable requests - 13.1%,
- unsatisfactory interest and assistance from user side - 12.4%,
- budget deficit, estimated budget, estimated finish time of parts developed software - 10.6%,
- no real expectations - 9.9%,
- insufficient support from management side of suppliers - 9.3%,
- change of requests and specifications - 8.7%,
- no successful planning - 8.1%,
- loss of need developed software - 7.5%.

He found out that the biggest problem is uncompleted or no understandable requests.

Nowadays there is an increasing request on development of software systems. Therefore flexibility, encapsulates, documentation, easy maintenance, easy configure, easy to implement new requirements, etc. are requirements on architecture of software systems. These next requests can be often in so large systems a matter of life and death.

A programmer and an analytic depend on level their own gained knowledge and experiences to try to reach these requirements during designing of developed software. From a programmer view it goes about transformations these

requirements to text models what represent source code, which can be directly compiled and run.

An analyst who has no knowledge about programming language is designing models from business view. In past was text form satisfactory, but today with large systems this method is no real to use, because of to keep understandable and readable knowledge about whole system. For this reason an analyst today are trying to use a visual-graphic level to transform requirements to the understandable models.

The paper is organized as follows. Section 1 introduces most used technologies for develop system in use with higher level of text (used standard programming languages) modeling. Section 2 introduces visual-graphic model based on UML tools. Sections 3 presents the conclusions.

## II. TEXT MODELING IN USE WITH STANDARD PROGRAMMING LANGUAGES

Today are using for modeling of software on source-code level technologies which are results of experts knowledge-base techniques e.g. patterns, spring, EJB, Hibernate, and an others. They cover knowledge of experts like are steps, repeated algorithms, etc. and they try to combine them to reach state of the platform independent, less failure an application, to reach faster development, maintenance. All these applied knowledge give a time frame to make or improve documentations of used architecture. We can say that each of these techniques is trying to get to higher level of an abstraction.

### A. Design Patterns

We can assume that ancestor of design patterns were objects. Patterns make them good with OOP community where experts-developers needed again and again apply same code steps. After time design patterns become to apply also in design, analyze and today we can to see patterns technique in different kind of and in more domain areas.

The pattern is helping to create OOP design, because they do identify classes, instances, their relationships, responsibility to solve concrete technical problems.

Design patterns solve many problems in various ways, with which a programmer meets daily. A team of Gang of Four[2] in their publication which is regards as bible of design patterns explains when, how and which concrete pattern to use.

Design patterns have also problems. One of them is dispersal of a pattern in a developed software system. There isn't still an acceptable way to get pattern to graphic form after a programmer use him. Also there is problem to identify used pattern in source code in developed software.

### B. Spring, EJB

During of developing more complex application a programmer meets with necessity to solve some ordinary aspects of software design e.g. authenticities, authorization, manage of transactions, permanent store data into database, etc. Next problem is weakness structure of application code, which is possible solve in use with a design pattern, but decomposition of wrote code is also uselessly repeatedly programming. These again and again repeated functions a Spring [3] framework (container) solves with their implementation into his (Spring) own classes so then the framework offers the whole infrastructure which developed applications are simply using. Property of a Spring framework is that take care about life cycle of an application and code is called by himself if it is needed.

Design pattern "*Inversion of Control*" gives possibility to move responsibility for create and co-operations of objects from application to a framework. This network is defined for example in a configure file (XML mainly).

EJB is very similar technology like Spring, but more complicated. Final application is dependent on concrete application server.

Main advantages of Spring:

- simplicity and directly using,
- independent on application environment,
- zero and minimal independent application code on Spring,
- all in one – Spring cover all layers of application - from present to persistent layer,
- unites principle of definition of objects and their relations in configure file, which allow for example take easier test of modules of an application.

### C. Hibernate

Hibernate technology[4] serve as objected relational mapping for possible easily using of relational databases with OOP languages. Persistence of objects then do not have to be solve in use with of particular files, XML etc., but data are directly mapped to (SQL) database.

ORM is technology depends on used platform.

Main advantages of ORM:

- faster and reliable saving of a great volume of data to relational databases,
- only objected oriented programming, source code is no fill up with large SQL queries,
- simply access to relational databases( is possible to divide development of DB a development of application).

Disadvantages of ORM:

- this technology is about mapping so it can lead to lower performance then native implementation,
- there is possible that couple constructions of queries can be harder achievable between object paradigm and relational DB.

### III. GRAPHIC – VISUALLY MODELING BASED ON UML TOOLS

In field of software engineering exists methodologies and approaches with which development is managed. Collective idea of these approaches is speed up and ensures better quality and maintenance of final software. A big motivation of these methodologies is also savings total cost on development. Methodologies cover complex techniques and steps how develop software.

One subset of methodologies is software design in use with computer graphic. This graphic implements improved transformation of text requests on development software and analytic is able to change, update and configure the software more complexly because of better view on architecture of system.

There are couple tools for modeling – designing requests in use with visual representation today (MDA, UML, DSM...)

The MDA[5] (Model Driven Architecture) technology is provided from OMG (Object Management Group). This group is focusing on making standards, which offers interoperability and portability of distributed OOP applications. Concept of MDA covers large part of exist specifications of OMG:

- UML (Unified Modeling Language),
- MOF (Meta-Object Facility),
- CWM(Common Warehouse Metamodel),
- XML (Extensible Markup Language),
- XMI (XML Metadata Interchange) a IDL (Interface Definition Language).

An idea of MDA is progressive specifying of models from higher layer of abstraction, which does include models of users without any relations to their implementations to lower layer, which contains models directly mapped to source code.

Some authors mark the UML in MDA as "UML as programming language", but just UML is not fully-fledged programming language.

Martin Fowler [6] defines three bases way of using UML:

- UML as sketch – UML is used only for catch of main ideas for developing software. This is main use of UML today,
- UML as blueprint – is trying to describe whole system detailed.
- Automated transformations of models to source code. These models have to contain many information, because of effective and complete transformations to source code,
- UML as programming language – system is completed described by models, so these diagrams became of run code so is possible to compile them to binary files.

Next advantage of MDA is possible to divide business logic from technology of platform. Result is that application made with MDA concepts are simply implemented in large scale of platforms (CORBA, J2EE, NET...). MDA allows develop an application on higher level of abstraction and wherefore the MDA left more time to focus on business logic and no need to take care about problems with implementation on a concrete platform.

MDA today already yield a no small advantages:

- created by implementation independent design improve transmittable of an application, what lead to saving of costs and reduction of development complexity for transform to another platform,
- By automatic generation is possible today to create almost 80% of source code,
- An automatic generation code does eliminate programmers – human mistakes.

On the other side the MDA technology has also opponents-critics. Martin Fowler said that UML did arise from “sketch” notation (as tools for capture important ideas and communication with programmer) but for use in the MDA is not the UML ready to use, yet.

Also he says that designing of sequence and activity diagrams is not better then to write code with modern programming language.

Steve Cook compares the MDA with a DSM (Domain Specific Modeling). DSM is next of possible approach to a develop software, where main artifact is a model. DSM doesn't try about automatic transform of models, instate of is based on creation of a model for each part of a system(domain) especially and after that verify or approve these models mutually.

Tools working on UML base with MDA implementation:

- *AndroMDA* –open-source product. Is using with another tools (ArgoUML, MagicDraw, Maven). Allows to write own transformational scripts also in JAVA, QVT, ALT.
- *Enterprise Architect (Sparx Systems)*, commercial product. Complete tools for designing, selecting of requests, etc. Covers all 13<sup>th</sup> of UML2.1 models.
- *OptimalJ* - commercial product. Contains model of processes based on activity diagrams in UML2.0. Supports Eclipse platform.
- *Borland Together 2006* - commercial product. Supports specification of (QVT) Query View Transformation, which allows run transfers among models. Supports also OCL2.0.
- *PowerDesigner9* – commercial product. Supports 8<sup>th</sup> implementations of MDA techniques,
- *Rational Rose* – commercial product. Supports couple implementations of MDA techniques.

#### IV. CONCLUSION

After compare of approach designing on text level and graphic level is evidently that both approaches are trying the complexity of designing software encapsulated to abstracted level and then make use of accessible knowledge.

Graphic representation of model is nevertheless understandable for business analytic and domain experts then text form. Interesting solution to graphic capture of information is concept of MDA even if there are also next similar concepts e.g. the DSM (Domain Specific Modeling).

The MDA concept is relative young technology, but if MDA success as fully-fledged programming language so

then it will be similar turn like in past the turn from assembler to higher programming languages.

The graphic form of designing is transformed to kind of a XML notation at the end and text form of designing is trying to manage run an application code also via any kind of XML structures so exist way how to connect graphic and textual form of designing and also to have better description of used software models. It can be reached during development when used programming language will demand to describe and determine relationships among another parts of system models.

Then we also to get state where programming language can tell what all parts of software system should be changed in the event when is need to change one part of system to keep integrity of whole system.

#### ACKNOWLEDGMENT

This work was supported by VEGA Grant No. 1/0350/08 Knowledge-Based Software Life Cycle and Architectures.

#### REFERENCES

- [1] S. L. Pfleeger, Design and analysis in software engineering: the language of case studies and formal experiments, in Plastics, New York, NY, USA, 1998.
- [2] E. Gamma, R. Heim, R. Johnson, J. Vlissides, Design Patterns, Holland 1998.
- [3] C. Walls, R. Breidenbach, Spring in Action, Manning, ISBN 1-933988-13-4, Greenwich, Aug.2007.
- [4] Ch. Bauer, G. King, Hiberante in Action, Mannin, ISBN 1932394-15-X, 209, Greenwich, 2006.
- [5] Cook, S., Domain-Specific Modeling and Model Driven Architecture, MDA Journal,2004.
- [6] Fowler, M., Model Driven Architecture, 2004.
- [7] Fowler, M., UML Distilled: A Brief Guide to the Standard Object Modeling Language, Addison-Wesley Professional, ISBN: 0321193687, 2003
- [8] Borland Software Corporation, Successful mplementation of Model Driven Architecture, 2007.
- [9] OMG 2nd Revised Submission: MOF2.0 QVT. <http://www.omg.org/cgi-bin/doc?ad/05-03-02>.
- [10] Kleppe, Anneke G., MDA Explained: The Model Driven Architecture: Practice and Promise, Addison-Wesley, ISBN:032119442X, 2003.
- [11] Kontio, M., Architectural manifesto: Choosing MDA tools, 2005.<http://www.ibm.com/developerworks/webservices/library/wi-arch18.html>
- [12] Gamma, E., Design Patterns: Elements of Reusable Object-Oriented Software, Addison-Wesley Professional, ISBN: 0201633612, 1994.



# Subgoal Discovery Methods in Reinforcement Learning

Marek Lapko

Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

marek.lapko@tuke.sk

**Abstract**—This paper is focused on algorithms that discover subgoals within reinforcement learning. The aim of discovering subgoals is to speed up learning by allowing transfer of knowledge on various levels of abstraction. To implement hierarchies, we consider option framework, as it allows transferring the knowledge even between tasks in various environments.

**Keywords**—q-learning, hierarchical reinforcement learning, subgoal finding, option

## I. INTRODUCTION

Reinforcement learning has proven to be a powerful tool in tasks, where (almost) all information we can provide learning agent is a goal to reach. The goal is determined by a reward, usually at the end of the episode. The aim of the agent is to learn by trial and error to get the highest reward along the path and thus reach the goal [1]. As the agent is provided just with a little information, it takes the agent a long time to learn, especially in the case when the agent is a real robot, not a simulation. One way, how we could speed up the learning, is to reuse the knowledge from previously learned tasks. To transfer the knowledge, we consider hierarchical reinforcement learning approach. The knowledge transfer in various levels of abstraction allows us to transfer and learn how to use even parts of knowledge in a new task. In this paper we focus on the problem, how to find a subgoals in a task. These subgoals divide the task into the parts, subtasks. Every such a task could be assigned to an option [2], which we can see as an action on the higher level of abstraction and which is easy to transfer [3][4].

In this paper, we focus on discovering subgoals, which we can see as a “funnels” or “bottle necks” in the state space [5]. In general, the bottle neck is a state (or a region of states), that lie on the path to the goal and it is critical to go through this state. There are several methods for finding such subgoals. In [6] as subgoals are determined the states with high frequency of visit or the states with a high reward gradient. Other approach to discover subgoals automatically is to use diverse density [5]. Krechmar et al. in [7] proposed algorithm that uses a combination of a frequency and a distance of a state from the start or from the end state. Subgoal discovered by algorithm in [8] is a state with a high gradient of a number of states leading to the state.

## II. REINFORCEMENT LEARNING

Reinforcement learning is a kind of learning, in which an agent, by interaction with an environment, tries to reach a goal. Usually, this problem is described as Markov Decision Process (MDP) [1]. An MDP is given by the tuple  $\langle S, A, T, R \rangle$ ,

where  $S$  is a set of states, that the agent can be in,  $A$  is a set of available actions,  $T : S \times A \times S \rightarrow [0, 1]$  is a transition function, which defines the probability of transition from the state  $s \in S$  to the state  $s' \in S$  when choosing action  $a \in A$  in state  $s$ .  $R : S \times A \rightarrow \mathcal{R}$  defines a reward, that the agent receives from the environment when choosing action  $a$  in state  $s$ . The goal is to find an optimal policy  $\pi : S \times A \rightarrow [0, 1]$ .

Q-learning is one of the most used RL algorithms. The update rule for Q-learning is defined as:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha \left[ r_{t+1} + \max_a Q(s_{t+1}, a) - Q(s_t, a_t) \right] \quad (1)$$

where  $Q(s_t, a_t)$  is the action-value function, which is a value of choosing action  $a$  in state  $s$  and following policy  $\pi$  and  $r_{t+1}$  is a actual reward given by the environment in time  $t + 1$ .

As our goal is to reuse the knowledge on various levels of abstraction, subgoal is represented by an option framework. Option is extended action - closed-looped policy for taking action over period of time [2]. Option is defined as a triple  $\langle I, \pi, \beta \rangle$ , where  $I \subseteq S$ , is an initiation set of states, states in which an option is available.  $\pi$  is the option’s policy and  $\beta(s)$  is the probability, that option will terminate in state  $s$ .

## III. SUBGOAL DISCOVERY METHODS

We implemented and compared two algorithms proposed by Goel in [8] and by Kretchmar in [7].

To find a subgoal by the first algorithm we have to define  $C : S \rightarrow \mathcal{N}$ , where  $S$  is a state set and  $\mathcal{N}$  is a set of natural numbers. In a deterministic environment,  $C(s)$  refers to the number of predecessors of a state  $s$  under given policy.  $C$  is defined recursively as:

$$C_1(s) = \sum_{s \neq s'} T(s|s', \pi(s')) \quad (2)$$

$$C_{t+1}(s) = \sum_{s \neq s'} T(s|s', \pi(s')) C_t(s') \quad (3)$$

$$C(s) = \sum_{i=1}^n C_i(s) \quad (4)$$

where  $T(s|s', \pi(s'))$  is the transition probability and  $n$  is the number of states or the algorithm stops recursion if  $C_{n+1} = C_n$ . When we have  $C$  for every state<sup>1</sup> we set the agent to a

<sup>1</sup>in real domain, it is almost impossible to have  $C$  for every state, but the more  $s$  the algorithm explored, the better results we get

“random” position. The agent, following given policy, reaches the goal. Along the trajectory, the agent did, we compute the gradient ratio

$$r_t = \frac{\Delta_t}{\Delta_{t+1}} \quad (5)$$

where

$$\Delta_t = C(s_t) - C(s_{t-1}) \quad (6)$$

Then, a state  $s_t$  with a ratio  $r_t$  higher than some threshold is considered to be a subgoal.

The second algorithm, designed in [7] is called FD algorithm. When finding subgoal by this algorithm, we consider the frequency of states as well as the minimal distance from the beginning or from the end of the gone trajectory. The algorithm consists of two steps:

- 1) to collect trajectories (we consider just trajectories that reaches the goal)
- 2) for every state we compute candidacy metric  $c_i$  (Eq. 7)

The candidacy metric is computed as:

$$c_i = F_i D_i \quad (7)$$

where  $F_i$  is the frequency of state (Eq. 8) and  $D_i$  is the distance metric (Eq. 9)

$$F_i = \frac{t_i}{T} \quad (8)$$

where  $t_i$  is the number of trajectories with state  $i$  and  $T$  is the number of all trajectories.

$$D_i = e^{-\left(\frac{1-d_i}{a}\right)^b} \quad (9)$$

where parameters were set to  $a = 0.5$  and  $b = 3.5$ .  $d_i$  is computed as:

$$d_i = 2 \min_{t \in T} \min_{s \in \{s_0, s_g\}} \frac{|s - i|}{l_t} \quad (10)$$

where  $s_0$  is a start state,  $s_g$  is a goal state and  $l_t$  is a trajectory length. Again, states with candidacy metric higher than some threshold are considered to be subgoals.

#### IV. EXPERIMENTAL RESULTS

All experiments were conducted in a simulated environment. The environment’s properties were set in the way to be similar to our LEGO robot environment.

Our environment has a planar rectangular shape, with two obstacles placed to create a “door” and to separate two spaces (Fig. 1). The environment is designed the way, that there is one obvious subgoal - the door.

We used Q-learning algorithm to find optimal policy with a table representation of knowledge and therefore we had to discretize the state space and the action space. The state space was represented by agent’s position in the environment, where the position was described by the  $x$  and  $y$  position and the agent’s orientation (an angle with the  $x$  axis). We had three discrete actions: turn right ( $+\alpha$ ), turn left ( $-\alpha$ ) and go straight.

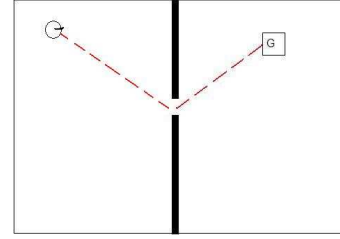


Fig. 1. Simulated environment. There is an agent (robot) in upper left corner. The door is a gap between two rectangular obstacles in the middle of the environment. In the upper right corner, there is a goal that the agent tries to reach. Red dashed line represent optimal trajectory.

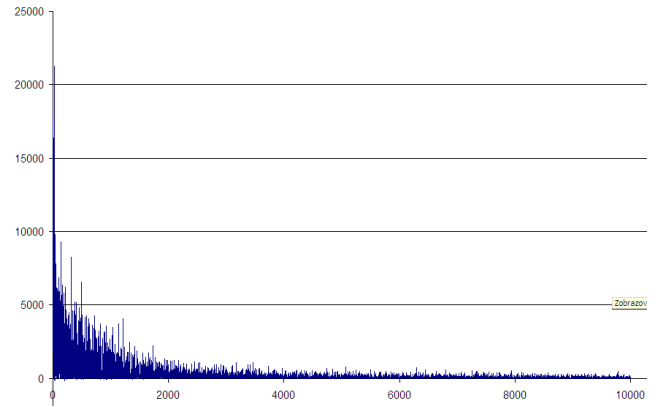


Fig. 2. Q-learning results. X - axis represents the number of episode and y - axis represents the number of cycles (actions) needed per episode to reach the goal.

##### A. Q-learning

Before discovering subgoals, the agent learned optimal policy of how to get from a random position to the goal. The size of the environment was  $600 \times 500$  with discretization step 20 units in both dimensions. The orientation of the agent was discretized by  $\frac{\pi}{2}$  step. The result was, that we had only four possible orientation (North, South, East, West). There were three possible actions for the agent turn  $\pm\frac{\pi}{2}$  and go straight 20 units. The reason for such discretization and action parameters was to keep the task MDP. Q-learning parameters were set to the values:  $\gamma = 0.9$ ,  $\epsilon \in [0.05, 0.01]$  and  $\alpha \in [0.1, 0.03]$ . The agent got reward  $+1.0$  when it reached the goal, otherwise it was punished every step with punishment  $-0.01$ .

As the problem, was designed to be MDP, the agent managed to reach the goal from every position optimally. In Fig. 2, we can see, that approximately after 8000 episodes the average number of cycles per episode was somewhere near 100 cycles.

##### B. Algorithm using count of predecessor states

In our case, we considered deterministic environment (MDP) and therefore the  $C$  values represent the count of the predecessor states. It means, that every value  $C(s)$  is equal to the number of states from which we can, by following given policy, reach the state. To find the subgoal, we placed the agent to a random position and computed the ratio (Eq. 5) along the trajectory. In (Fig. 3) there is depicted a ratio  $r$  along one trajectory. The highest  $C$  value we have got in the state, where the agent is in front of the door and is facing the door (Fig. 4). Which mean, that the algorithm found the subgoal at the same position we presumed.

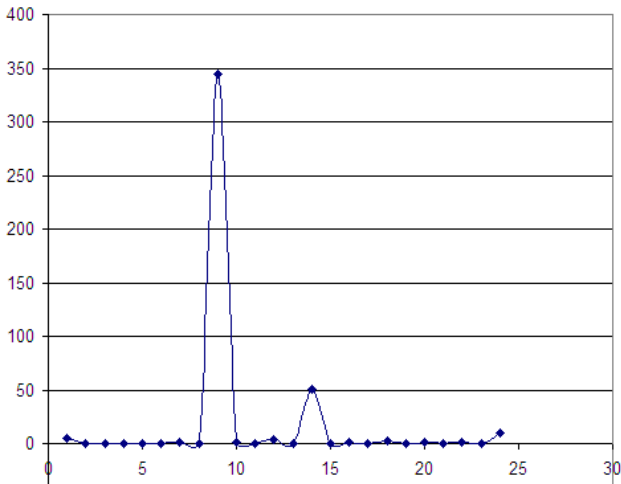


Fig. 3. Ratio along the trajectory. X-axis represents time along the trajectory and y-axis represents ratio value. There is one significant peak, which represents subgoal.

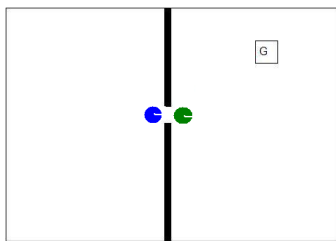


Fig. 4. Subgoal state represented in the environment. The blue agent is in the position of the subgoal found by the algorithm using count of predecessor states and the green agent is in the position of the subgoal found by the FD algorithm

This algorithm works well if we have an optimal policy for the whole state space and the problem is MDP. Otherwise we have to skip such states from computing  $C$  that are part of loops.

### C. FD algorithm

Similarly, as in previous algorithm, we had found optimal policy before we applied algorithm to discover the subgoal. Then, to compute  $c$  values for all states, algorithm explored approximately 500 trajectories with random start state. Results are depicted in Fig. 5.

In this case, if we are looking just for one subgoal, the subgoal is evident. But as we can see in the figure, there are more states with very high candidacy metric. There seems to be very simple explanation for this situation, because all this states lie on the line that leads to the subgoal (goal) The line is collinear with axis because we allowed the agent to be in just four possible orientations  $(0, \frac{\pi}{2}, \pi, \frac{3\pi}{2})$ . Another interesting thing is that this algorithm found not the same state as a subgoal as the previous algorithm did. The subgoal state according to this algorithm is right behind the door.

## V. CONCLUSIONS AND FEATURE WORK

In this paper, we have demonstrated two algorithms for subgoal discovering within reinforcement learning. Both algorithms found subgoals and even though these subgoals were not the same state, both were near the state we considered to be a subgoal.

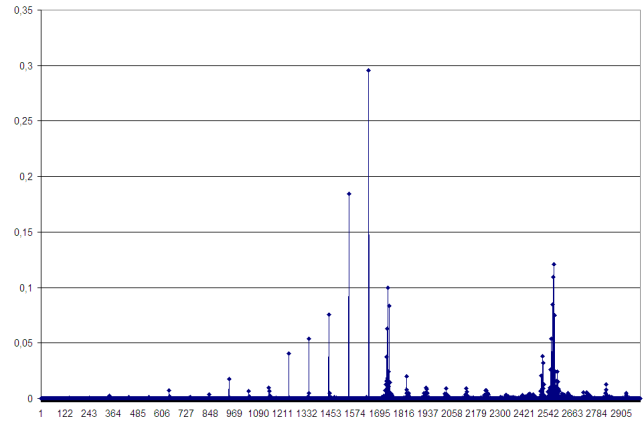


Fig. 5.  $c$  values for every state from FD algorithm. X-axis represents the state and y-axis represents the  $c$  value.

In the future work, we would like to focus on finding subgoals on real domains like LEGO NXT where subgoals do not have to be so evident and on building portable options. Another challenge is to automatically determine which state, according to the  $c$  value is a valuable subgoal. It means, which local maximum we can consider being a subgoal and which is just a common state.

## REFERENCES

- [1] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction (Adaptive Computation and Machine Learning)*. The MIT Press, March 1998.
- [2] R. Sutton, D. Precup, and S. Singh, "Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning," *Artificial intelligence*, vol. 112, no. 1, pp. 181–211, 1999.
- [3] M. Lapko and R. Jakša, "Reuse of knowledge within reinforcement learning in mobile robotics," *13th International Conference on Cognitive and Neural Systems 2009 (accepted)*, 2009.
- [4] M. Lapko, "Reuse of knowledge in reinforcement learning (in Slovak)," *Thesis proposal*.
- [5] A. McGovern and A. Barto, "Automatic discovery of subgoals in reinforcement learning using diverse density," pp. 361–368, 2001.
- [6] B. Digney, "Learning hierarchical control structures for multiple tasks and changing environments," p. 321, 1998.
- [7] R. Kretschmar, T. Feil, and R. Bansal, "Improved automatic discovery of subgoals for options in hierarchical reinforcement learning," *Journal of Computer Science and Technology*, vol. 3, no. 2, pp. 9–14, 2003.
- [8] S. Goel and M. Huber, "Subgoal Discovery for Hierarchical Reinforcement Learning Using Learned Policies," 2003.

# Finite-State Machine in Speech Recognition

<sup>1</sup>Martin Lojka, <sup>2</sup>Marek Papco

Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

<sup>1</sup>martin.lojka@tuke.sk, <sup>2</sup>marek.papco@tuke.sk

**Abstract**—Most modern automatic speech recognition systems (ASRs) are using statistical information that is extracted from large amount of data sets. This statistical information is represented by acoustic and language model. Acoustic model is mostly based on HMM and language model on n-gram stochastic model. The element between acoustic model and language model is lexicon, which contains word to phonemes transcription.

These model are obtained by using modern learning techniques and given in form of finite-state machine (FSM) directly or as an approximation of more complex models. In order to search for the most probable word sequence (for continuous speech recognition) a recognition network from FSM is created.

This paper provides an overview to the speech recognition using FSM for the construction of a recognition network. In first place we will look at the speech recognition process and then at the most significant operations for the construction of a recognition network.

**Keywords**—Speech recognition, Finite state acceptors, Finite state transducers, Recognition network

## I. INTRODUCTION

The basic idea of stochastic system for automatic speech recognitions (ASR) is to use a statistical information about structure of sounds, words and sentences in particular language. The ASR system can be broken into two main parts. First part is extracting observation vectors from input speech. In that way the data rate to the next, the most time-consuming part, can be lowered. The second part is processing all provided statistical information about language and the acoustic observation vectors in order to "decode" input speech. We use word "decode" because the user, as he speaks, is "encoding" the words from his mind into sounds coming from his mouth. On the side of ASR system we are then performing a reverse operation, the "decoding".

The decoding part of speech recognition can be formulated as follows [1]:

$$\hat{W} = \arg \max_W [P(W|O)] \quad (1)$$

From formula (1) is clear that the speech recognition problem can be explained as finding the most probable sequence of words. Since  $P(W|O)$  is difficult to model directly, Bayes' rule is used to transform the (1) into the following formula:

$$\hat{W} = \arg \max_W [P(O|W)P(W)] \quad (2)$$

The formula (2) is showing us that modeling problem that is required for speech recognition can be broken into two parts.

First part is model that provides a estimation of  $P(O|W)$  probability. The model is usually based on Hidden Markov Models (HMM) and contains statistical information for classification of sounds to particular units of speech. The most used speech units in acoustic modeling are phones and phones with

respect to neighbor phones in speech (the context-dependent phones) – triphones.

The second part is probability  $P(W)$  that is provided by language model. Language model is useful in concatenating recognized words in to the sentences. The most used are n-gram language models, which they can provide the probability of particular word based on its history. For that purpose trigrams (two word history is considered) or bigrams (one word history is considered) are used.

The one missing part in this structure of ASR system is the way how to bind these two models. The solution is to create a pronunciation lexicon, which contains a list of words (the whole vocabulary for ASR system) with a pronunciation (the sequence of phones).

As was mentioned earlier, the decoding process of speech can be seen as problem of finding the most probable sequence of words. With a combination of knowledge sources a recognition network can be created where the most probably path in this network defines the recognized word sequence. There are number of methods that can be used for finding the right path. The most used methods are *Viterbi* decoding [2][3] using *beam* search [2][3][4][5] or *A-stack* [6] method, which is a variant of *A\** [7] algorithm. There are also decoders that are using combination of both methods [8].

By combining all knowledge sources a recognition network can be created at the beginning of decoding process or dynamically during the process. Networks created before decoding process are called *static* [9] and have one major disadvantage, which is the size of the network for large vocabulary continuous speech recognition (LVCSR). This can be partially managed by introducing minimization techniques into the network construction and take an advantage of redundancies.

The recognition network is based on finite-state acceptor (FSA) or finite-state transducer (FST). The FSA is basically the same as FSM, but in this paper we will use the notation FSM if we are speaking about FST and FSA generally.

## II. RECOGNITION NETWORK

For the search process for the most probable word sequence, which is recognized sentence, a search space is needed. This search space is based on search network, which is created from FSM. As for recognition network the FSMs can be used. There are two basic types of the FSMs. First one is FSA and the second one FST.

As shown in the Fig.1 and Fig.2 the FSA and FST are represented by a finite set of states, from which one is initialization state and one or more final states. States are connected with transitions, which are labeled with symbols (FSA has one label, FST two labels) and are weighted. In

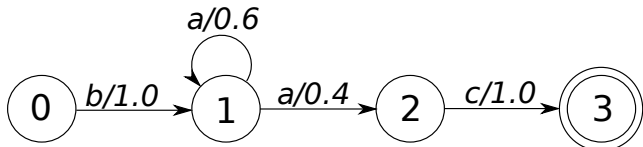


Fig. 1. Example of a finite state acceptor (FSA) with weighted transitions

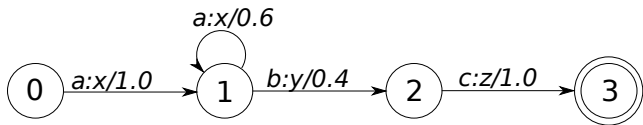


Fig. 2. Example of a finite state transducer (FST) with weighted transitions



Fig. 3. Recognition cascade

this publication we will be using notation "symbol/weight" for labeling transitions. For FST we will use notation "input symbol:output symbol/weight". The weights are probabilities that were obtained by learning techniques from large data set.

The network construction with FSA is based on simple substitution technique [10]. For example the word in grammar FSA is replaced by the pronunciation transcription from lexicon FSA. For FST we use recognition cascade as shown in the Fig. 3 [11].

As shown in the Fig. 3 the FST provides mapping from one level of representation to another. For example the context-dependent HMM FST "C" provides mapping of context-dependent phones to context independent. In this principle the mapping from input observations to sequences of recognized words can be done. FST "H" provides mapping from distribution functions of HMM models to context-dependent phones. FST "L" is mapping context-independent phones to words. FST "G" is mapping from word to words, so it has the same output and input labels.

In the rest of this paper we will discuss operations that are used for construction of a recognition network with FST. In some cases we will demonstrate the operations on FSA for simplicity. Before the operations, we need to say what is a *semiring*. The operations and FSMs are dependent of the type of semiring used.

### III. SEMIRING

The speech recognition using FSM depends on the path through FSM. The path defines the recognized words. In most of the cases during speech recognition there are multiple path from initialization state to one of the final states. In this case we must know how to combine weights along one path and resulting weights of all paths. Thus we must have a method to know which word sequence is the most probable. The way, in which we handle the weights, is called semiring [10][11]. Semiring is defined as  $(\mathbf{K}, \oplus, \otimes, \bar{0}, \bar{1})$ , specifically by a set of values  $\mathbf{K}$ , two binary operations  $\oplus$  and  $\otimes$ , and two designated values  $\bar{0}$  and  $\bar{1}$ . The operations  $\oplus$  is associative, cumulative, and has  $\bar{0}$  as identity. The operation  $\otimes$  is associative, has identity  $\bar{1}$ , distributes with respect to  $\oplus$ , and has  $\bar{0}$  as annihilator (3).

$$\bar{0} \otimes a = a \otimes \bar{0} = \bar{0} \quad (3)$$

 TABLE I  
SEMIRING EXAMPLES

SEMIRING	SET	$\oplus$	$\otimes$	$\bar{0}$	$\bar{1}$
Probability	$R_+$	+	$\times$	0	1
Log	$R \cup -\infty, +\infty$	$\oplus_{\log}$	+	$+\infty$	0
Tropical	$R \cup -\infty, +\infty$	min	+	$+\infty$	0

$$x \oplus_{\log} y = -\log(e^{-x} + e^{-y})$$

Here are few possibilities, everyone with his advantages and disadvantages (Table I).

The probability operator (Table I)  $\otimes$  is used for combination of probabilities in one path and  $\oplus$  for combination of resulting probabilities of different paths. The most used is tropical semiring that is derived from the log semiring using *Viterbi approximation*.

### IV. STRUCTURE OF A FST

Language is a sequential phenomena. Words occur in sequence over time. Words that appeared before are used for constraining the interpretation of the following words. For the simplest models that are capable to model such a sequential phenomena FST or FSA can be used.

In Fig. 1 is an example of FSA and in Fig. 2 an example of FST. Both have weighted transitions. Thus we can call them weighted FSA (WFSA) and weighted FST (WFST).

WFST  $T = (\Sigma, \Omega, Q, E, i, F, \lambda, \rho)$  over semiring  $\mathbf{K}$ , is specified by a finite input alphabet  $\Sigma$ , a finite output alphabet  $\Omega$ , a finite set of states  $Q$ , a finite set of transitions  $E \subseteq Q \times \Sigma \times (\Omega^+ \cup \{\epsilon\}) \times \mathbf{K} \times Q$ , an initial state  $i \in Q$ , a finite set of final states  $F$ , an initial state weight assignment  $\lambda$  and a final state weight assignment  $\rho$  [12][13].

The given transition  $t = (p[t], l_I[t], l_O[t], w[t], n[t]) \in E$  is specified by a previous state or origin of the transition  $p[t] \in Q$ , a next or destination state  $n[t] \in Q$ , its weight of the transition  $w[t]$ , its input label  $l_I[t]$  and its output label  $l_O[t]$ . Every path through WFST  $T$  is sequence of transitions  $t_1, t_2, \dots, t_n$  while  $n[t_i] = p[t_{i+1}]$ ,  $i = 1, 2, \dots, n - 1$ . The successful path  $\pi = t_1, t_2, \dots, t_n$  is a path from initial state  $i$  to the one of the final states  $f \in F$ . The resulting string from this transducer is the concatenation of all label that are associated with transitions  $(l_I[\pi], l_O[\pi]) = (l_I[t_1], l_O[t_1])(l_I[t_2], l_O[t_2]) \dots (l_I[t_n], l_O[t_n])$ . The resulting weight for one successful path through transducer is a  $\otimes$ -product of the weights along this path  $\pi$ ,  $w[\pi] = \lambda \otimes w[t_1] \otimes \dots \otimes w[t_n] \otimes \rho(n[t_n])$ . The total weight for one input sequence is  $\oplus$ -product of all successful paths.

WFSA is defined in similar way by simply omitting input or output labels. Also can be said that WFSA is WFST with the same input and output labels.

### V. OPERATIONS WITH FSTs

Using operations with finite state transducer the recognition network can be created. For FST and also for FSA are defined many operations (Table II) [14]. In this paper we will provide an overview of the most significant ones.

#### A. Composition

The composition operations with FST is the most important operation in area of speech recognition. This operation allows

TABLE II  
 OPERATIONS WITH FST

Operation	Type	Usage
Composition	Binary transform	Knowledge source combination
Concatenation	Binary transform	Sequencing
Connection	Lossless optimization	Redundancy removal
Determinization	Lossless optimization	redundancy removal
Difference	Binary transform	Model restriction
Epsilon removal	Lossless optimization	Model symmetry
Intersection	Binary transform	Model restriction
Inversion	Unary transform	Input/output inversion
Kleene Closure	Unary transform	Indefinite repetition
Minimization	Lossless optimization	Space optimization
Projection	Unary transform	Output removal
Pruning	Lossy optimization	Search optimization
Reversal	Unary transform	Sequence reversal
Shortest path	Lossy optimization	Search optimization
Union	Binary transform	Model merging
Weight pushing	Lossless optimization	Search optimization

us to combine two FSTs of knowledge sources as shown in Fig. 3. Every FST represents information about knowledge sources like lexicon, grammar or language model, and acoustic model. Consider two FSTs  $T_a$  and  $T_b$  as shown in the Fig. 4 and Fig. 5 and defined by (4). The  $T_a$  provides mapping from all sequences  $\Sigma_a^*$  to output sequences  $\Omega_a^*$ . Where  $\Sigma_a^*$  and represents set of all input sequences that can be constructed from symbols in alphabet  $\Sigma_a$ . The same for output sequences represents notation  $\Omega_a^*$ . The next FST  $T_b$  in recognition cascade provides further mapping from  $\Omega_a^*$  to  $\Omega_b^*$ . This also means that the input alphabet  $\Sigma_b$  of  $T_b$  must be the same as output alphabet  $\Omega_a$  of  $T_a$ , thus  $\Sigma_b = \Omega_a$ . This mapping can be done in one step by composition of FSTs  $T_a$  and  $T_b$ . The resulting FST is defined by (5)[14].

$$\begin{aligned} T_a &= (\Sigma_a, \Omega_a, Q_a, E_a, i_a, F_a, \lambda_a, \rho_a) \\ T_b &= (\Sigma_b, \Omega_b, Q_b, E_b, i_b, F_b, \lambda_b, \rho_b) \end{aligned} \quad (4)$$

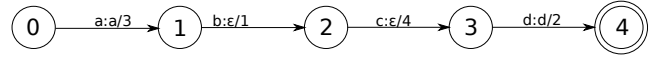
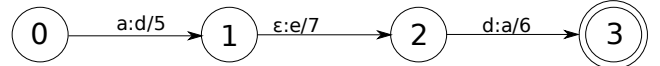
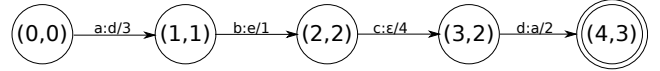
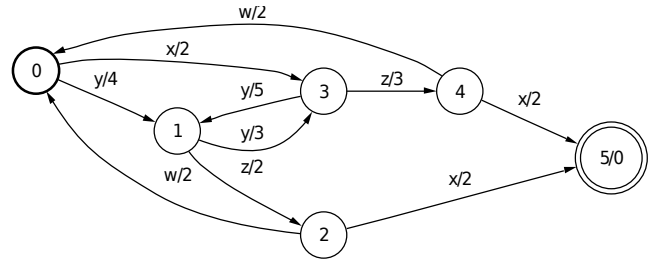
$$T_c = (\Sigma_a, \Omega_b, Q, E, i, F, \lambda_b, \rho_b). \quad (5)$$

Let the writing  $[T_a](\alpha \in \{\Sigma_a^*\}, \beta \in \{\Omega_a^*\})$  and  $[T_b](\alpha \in \{\Sigma_b^*\}, \beta \in \{\Omega_b^*\})$  represent the mapping of FSTs  $T_a$  and  $T_b$ , the function of composition of two FSTs is defined by (6). The notation for this operation is  $\circ$ , thus  $T_c = (T_a \circ T_b)$ . The resulting weight of transitions are a  $\otimes$ -product of particular weights of original two transducers.

$$[T_c](\alpha, \beta) = [T_a \circ T_b](\alpha, \beta) = \bigoplus_{\gamma} [T_a](\alpha, \gamma) \otimes [T_b](\gamma, \beta)$$

The basic principle of composition can be summarized as follows [10]:

- 1) The initial state of the  $T_c$  is a pair of the initial states of the  $T_a$  and the  $T_b$ .
- 2) The final state of the  $T_c$  is a pair of final states of the  $T_a$  and the  $T_b$ .
- 3) In the resulting FST  $T_c$  there is a transition from pair of states  $(q_1, q_2)$  to  $(r_1, r_2)$  for each transition in  $T_a$  from  $q_1$  to  $r_1$  and in  $T_b$  from  $q_2$  to  $r_2$ , where the output symbol in  $T_a$  is the same as in  $T_b$ . The resulting transition is


 Fig. 4. Transducer  $T_a[10]$ 

 Fig. 5. Transducer  $T_b[10]$ 

 Fig. 6. Resulting transducer  $T_c = T_a \circ T_b$ 

 Fig. 7. Acceptor  $A_a[14]$ 

then labeled with input symbol from the transition in  $T_a$  and output symbol from the transition in  $T_b$ . The resulting weight is the  $\otimes$ -product of particular weights. On the Fig. 4, Fig. 5 are FSTs that are about to compose and on Fig. 6 is the resulting FST.

### B. Weight Pushing

This operation is needed before minimization operation of the transducer. The resulting transducer from this operations has "pushed" weights towards initial state. This operations has also his advantage in applying the weights as soon as possible for effective pruning during the search recognition process.

Let  $A_a$  be the acceptor over semiring  $\mathbf{K}$  then we assume that this summation (6) exists.

$$d[q] = \bigoplus_{\pi \in P(q, F)} (w[\pi] \otimes \rho[n[\pi]]) \quad (6)$$

where the  $d[q]$  is the shortest path from  $q$  to one of the final states  $F$ . The operation of pushing weights is about recomputing the weights of initial, final and all transitions in the following way (7) [10]. The example of weight pushing of FSA  $A_a$  on Fig. 7 is on Fig. 8.

$$\begin{aligned} \lambda[i] &\leftarrow \lambda[i] \otimes d[i] \\ \rho[f] &\leftarrow d[f]^{-1} \otimes \rho[f] \text{ if } d[f] \neq \bar{0} \\ w[e] &\leftarrow d[p[e]]^{-1} \otimes w[e] \otimes d[n[e]] \text{ if } d[p[e]] \neq \bar{0} \end{aligned} \quad (7)$$

### C. Minimization

The minimization operation can be applied to deterministic FST or FSA[15]. In this operation we can join *equivalent* states and in this way we can get transducer with less states and transitions. Two states of FST or FSA are equivalent if the path to final state is labeled with the same symbols and weights

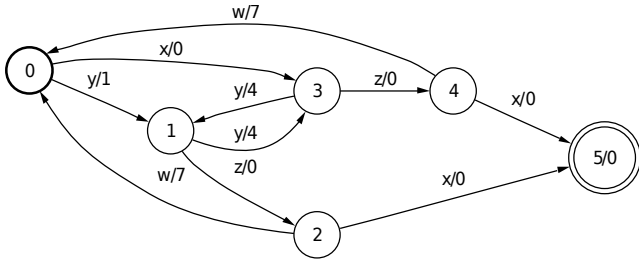


Fig. 8. Acceptor with pushed weights  $A_b$

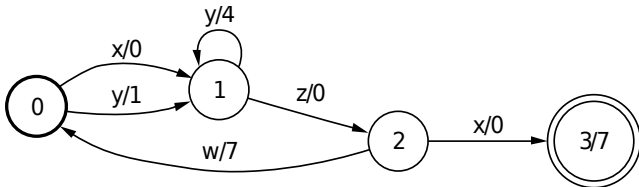


Fig. 9. Acceptor from minimization process of transducer from Fig. 8[14]

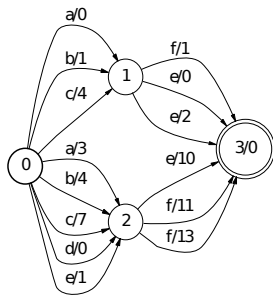


Fig. 10. Non-deterministic transducer[10]

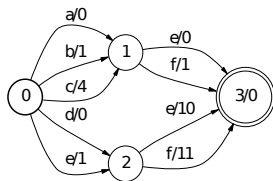


Fig. 11. Deterministic transducer[10]

of this path including weight of the final state are the same. Equivalent states can be joined without destroying the function of this FST or FSA. In real case there aren't such equivalent states, but with weight pushing operation we can create them and so we can apply the minimization operation. On the Fig. 9 is the example of minimization of the FSA from Fig. 8.

For acceptor we can use only weight pushing, but for transducer we need to "push" also output symbols to achieve path with the same labeling.

**D. Determinization**

The FSA or FST are deterministic if there is one unique initial state and no two transition leaving any state share the same input label [16]. If the FSA or FST is deterministic, the input sequence exactly determines the output sequence.

**ACKNOWLEDGMENT**

The work presented in this paper was supported by the Ministry of education of Slovak Republic under research projects AV 4/0006/07 and AV 4/2016/08 and Slovak Research and Development Agency under research project APVV-0369-07

**VI. CONCLUSION**

The main principle of speech recognition, as described in the introduction section, remains the same for FST and FSA. For the FST the decoder can be simpler and more general, which is advantage of the FST. Using FST, the RN is no longer dependent of the implementation of the decoder. Described operations in this paper are used for constructing final recognition network, which can be static or can be created dynamically during speech recognition process.

**REFERENCES**

- [1] X. Huang, A. Acero, and H.-W. Hon, *Spoken Language Processing: A Guide to Theory, Algorithm, and System Development*. Upper Saddle River, NJ, USA: Prentice Hall PTR, 2001.
- [2] L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," in *Proceedings of the IEEE*, 1989, pp. 257–286.
- [3] S. Patel, "An o(n/e) viterbi algorithm," in *ICASSP '97: Proceedings of the 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '97)-Volume 3*. Washington, DC, USA: IEEE Computer Society, 1997, p. 1795.
- [4] I. Trancoso, "A decoder for finite-state structured search spaces," in *ASR 2000 Workshop*, 2000, pp. 18–20.
- [5] H. Ney and S. Ortmanns, "Dynamic programming search for continuous speech recognition," *IEEE Signal Processing Magazine*, 1999.
- [6] D. B. Paul, "An efficient a\* stack decoder algorithm for continuous speech recognition with a stochastic language model," in *HLT '91: Proceedings of the workshop on Speech and Natural Language*. Morristown, NJ, USA: Association for Computational Linguistics, 1992, pp. 405–409.
- [7] —, "Algorithms for an optimal a\* search and linearizing the search in the stack decoder," in *ICASSP '91: Proceedings of the Acoustics, Speech, and Signal Processing, 1991. ICASSP-91., 1991 International Conference*. Washington, DC, USA: IEEE Computer Society, 1991, pp. 693–696.
- [8] P. S. Gopalakrishnan, L. R. Bahl, and R. L. Mercer, "A tree search strategy for large-vocabulary continuous speech recognition," in *Acoustics, Speech, and Signal Processing, 1995. ICASSP-95., 1995 International Conference on*, vol. 1, 1995, pp. 572–575 vol.1.
- [9] X. L. Aubert, "A brief overview of decoding techniques for large vocabulary continuous speech recognition," in *ASR-2000*.
- [10] F. C. N. P. Mehryar Mohri and M. Riley, "Speech recognition with weighted finite-state transducers," *Handbook on Speech Processing and Speech Communication*, 2008.
- [11] M. Mohri, *Statistical Natural Language Processing*. Cambridge University Press, 2005.
- [12] F. C. N. P. Mehryar Mohri and M. Riley, "Weighted finite-state transducers in speech recognition," *Computer Speech and Language*, 2002.
- [13] J. Psutka, L. Müller, J. Matoušek, and V. Radová, *Mluvíme s počítačem česky*. Prague: Academia, 2006.
- [14] A. Seward, "Efficient methods for automatic speech recognition," Ph.D. dissertation, Royal Institute of Technology Stockholm, 2003.
- [15] M. Mohri, "Minimization algorithms for sequential transducers," *Theoretical Computer Science*, 2000.
- [16] M. Mohri and M. Riley, "Network optimizations for large vocabulary speech recognition," *Speech Communication*, 1999.

# Design and Application of Optimal Control of Education Model Ball & Plate

Richard LONŠČÁK

Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

richard.lonscak@tuke.sk

**Abstract**—The paper is focused on optimal control design with application for solving problem tracking changes of the references trajectories by simulation model of mechanical system - Ball & Plate. The modified optimal control algorithm based on LQ principle with integral action use linearized model of the dynamic system in the state space and it's verified by designed simulation schemes in language Matlab/Simulink using the architecture of S-functions. The paper presents results of the tracking of the reference trajectories by nonlinear model of mechanical system using designed optimal LQ algorithm with integral action too.

**Keywords**— optimal control, nonlinear system, state space model, cost function, minimization of quadratic criterion, simulation.

## I. INTRODUCTION

The goal of this paper is the optimal control algorithm design with integral action and it's verification on the simulation model Ball & Plate [2].

The optimal control algorithm based on LQ principle – LQ control design with integral action (in discrete time domain – LQ algorithm with summator) [6] presented in this paper uses mathematical-physical nonlinear model of the dynamic system in the state space, which is linearized at the set point near to the steady state.

The designed modified algorithm of the optimal control is verified by simulation control schemes in language Matlab/Simulink on simulation model of the mechanical system Ball & Plate.

## II. OPTIMAL LQ CONTROL – CLASSICAL APPROACH

The problem of the synthesis of the classical approach to the optimal control into an equilibrium state for the controlled discrete linear system, which is defined by the state space model (1)

$$\begin{aligned} \mathbf{x}(k+1) &= \mathbf{A}\mathbf{x}(k) + \mathbf{B}\mathbf{u}(k) \\ \mathbf{y}(k) &= \mathbf{C}\mathbf{x}(k) + \mathbf{D}\mathbf{u}(k) \end{aligned} \quad (1)$$

where

$\mathbf{x}(k)$  is the state vector, dimension  $n$ ,

$\mathbf{u}(k)$  is the vector of the control inputs, dimension  $r$ ,

$\mathbf{y}(k)$  is the output vector, dimension  $m$ ,

$\mathbf{A}$  is the matrix of the dynamics ( $n \times n$ ),  
 $\mathbf{B}$  is the matrix of the control ( $n \times r$ ),  
 $\mathbf{C}$  is the matrix of the output ( $m \times n$ ),  
 $\mathbf{D}$  is the output matrix ( $m \times r$ ), in generally  $\mathbf{D} = 0$

is to determine such design of the algorithm of the control

$$\mathbf{u}(k) = -\mathbf{K}(k)\mathbf{x}(k) \quad (2)$$

which minimize of the quadratic cost function at the control during final interval of the time

$$J_{LQ} = \sum_{k=0}^{N-1} [\mathbf{x}^T(k)\mathbf{Q}\mathbf{x}(k) + \mathbf{u}^T(k)\mathbf{R}\mathbf{u}(k)] + \mathbf{x}^T(N)\mathbf{Q}^*\mathbf{x}(N) \quad (3)$$

while the matrices of the weights in the cost function  $J_{LQ}$  are determined so, that  $\mathbf{Q}^*$  and  $\mathbf{Q}$  are the positive semidefinite matrices of the size  $n \times n$  and  $\mathbf{R}$  is the positive definite matrix of the size  $r \times r$ .

The matrix  $\mathbf{K}(k)$  in the control law (2) is the matrix of the feedback gain of the size  $r \times n$ , ( $n$  is the number of the state of the system,  $r$  is the number of the inputs to the system).

Because at the real systems the control input  $\mathbf{u}(k)$  is always constrained, the task of the matrix  $\mathbf{R}$  in the cost function (3) is to ensure the constraint of the elements of the control vector  $\mathbf{u}(k)$  to the physical feasible values. The meaning of the positive semidefinite matrix  $\mathbf{Q}$  at the control is to secure of the convergence of the elements of the state vector to zero. The positive semidefinite matrix  $\mathbf{Q}^*$  represents of the weight of the vector of the state variables  $\mathbf{x}(N)$  in the step  $N$ , when the vector of the control variables is equals zero yet, t.j.  $\mathbf{u}(N) = 0$  and the matrix  $\mathbf{Q}^*$  we can choose so, that  $\mathbf{Q}^* = \mathbf{Q}$ .

For dynamic system, which is defined in the state space (1) we can calculate an optimal of the sequence of the symmetrical positive definite matrices  $\{\mathbf{P}(k), k = N-1, N-2, \dots, 0\}$ , which satisfy of the boundary conditions  $\mathbf{P}(N-1) = \mathbf{Q}^*$ ,  $\mathbf{P}(-1) = \mathbf{P}(0)$  and can be obtained by solving of the discrete Riccati equations



$$\begin{aligned} \mathbf{P}(k-1) = & \mathbf{Q} + \mathbf{A}^T \mathbf{P}(k) \mathbf{A} - \\ & - \mathbf{A}^T \mathbf{P}(k) \mathbf{B} (\mathbf{R} + \mathbf{B}^T \mathbf{P}(k) \mathbf{B})^{-1} \mathbf{B}^T \mathbf{P}(k) \mathbf{A} \end{aligned} \quad (4)$$

From this sequence we can determine desired sequence of the matrices of the feedback gain

$$\{\mathbf{K}(k), k = N-1, N-2, \dots, 0\}$$

from the equation

$$\mathbf{K}(k) = (\mathbf{R} + \mathbf{B}^T \mathbf{P}(k) \mathbf{B})^{-1} \mathbf{B}^T \mathbf{P}(k) \mathbf{A} \quad (5)$$

The condition of the existence of the solution for control to the equilibrium state is the controlled dynamic system (1). The acquired values of the matrix of the feedback gains  $\mathbf{K}(k)$  for defined control interval we can use for the calculation of the control signal by (2) for  $k = 0, 1, 2, \dots, N-1$ . By this means calculated control law is called **Linear Quadratic Control – LQ Control** [1], [7].

### III. MODIFIED LQ CONTROL ALGORITHM

The goal of this part is optimal LQ control algorithm design with summator with the application for solving problem of tracking changes of the reference trajectories without steady state error by simulation Ball&Plate model [2] and for solving compensation problem of the permanent disturbances actuating to the input/output of the system in the control structure on Fig.1.

The modified control algorithm based on LQ principle uses linearized model at the set point near to the steady state of the nonlinear system in the state space and the feedback gain control matrix  $\mathbf{K}$  of the state controller, which is computed by (4) and (5) minimization of the cost function (3).

For optimal control algorithm design with summator [6] we shall consider the extended state description of the closed-loop, which we get possession of by arrangement of the control signal  $\mathbf{u}(k+1)$ , which is equals:

$$\mathbf{u}(k+1) = -\mathbf{K}\mathbf{x}(k+1) = -\mathbf{K}[\mathbf{A}\mathbf{x}(k) + \mathbf{B}\mathbf{u}(k)] \quad (6)$$

where  $\mathbf{K}$  is the feedback gain matrix of the state controller (computed using the classical algorithm of LQ control) and ensures optimal compensation of non-zero vector of an initial values  $\mathbf{x}(0)$  [7].

The extended state description of the closed-loop is equal

$$\begin{bmatrix} \mathbf{x}(k+1) \\ \mathbf{u}(k+1) \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ -\mathbf{KA} & -\mathbf{KB} \end{bmatrix} \begin{bmatrix} \mathbf{x}(k) \\ \mathbf{u}(k) \end{bmatrix} \quad (7)$$

We shall consider the dynamic system in the state space with permanent disturbances actuating to the input/output of the system in modified control structure (Fig. 1) in the form:

$$\begin{aligned} \mathbf{x}(k+1) &= \mathbf{A}\mathbf{x}(k) + \mathbf{B}\mathbf{u}(k) + \mathbf{B}_d \mathbf{z}(k) = \\ &= \mathbf{A}\mathbf{x}(k) + \mathbf{B}\mathbf{u}(k) + \mathbf{d} \end{aligned} \quad (8)$$

where for unmeasured value of the disturbances hold  $\mathbf{d} = \mathbf{B}_d \mathbf{z}(k)$ .

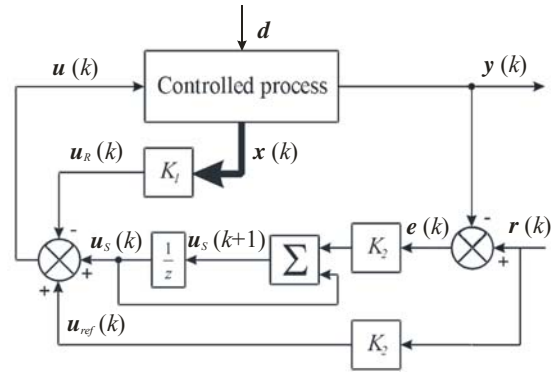


Fig. 1 Modified control structure with summator and forward loop with gain  $\mathbf{K}_2$

To be able to ensure tracking of the changes of the reference trajectory we shall consider modified control structure with summator (Fig.1), from which implies that the control input to the dynamic system (8) is created by three elements:

$$\begin{aligned} \mathbf{u}(k) &= \mathbf{u}_R(k) + \mathbf{u}_S(k) + \mathbf{u}_{ref}(k) = \\ &= -\mathbf{K}_1 \mathbf{x}(k) + \mathbf{u}_S(k) + \mathbf{K}_2 \mathbf{r}(k) \end{aligned} \quad (9)$$

Such defined control input for step  $(k+1)$  and the state of the dynamic system with disturbances (8) substituted to the extended state space description of the closed loop (7) and we get

$$\begin{aligned} \begin{bmatrix} \mathbf{x}(k+1) \\ \mathbf{u}(k+1) \end{bmatrix} &= \begin{bmatrix} \mathbf{0} \\ \mathbf{K}_2 \end{bmatrix} \mathbf{r}(k+1) + \begin{bmatrix} \mathbf{I} \\ -\mathbf{K}_1 \end{bmatrix} \mathbf{d} + \\ &+ \left\{ \mathbf{I} + \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{K}_1 & -\mathbf{K}_2 \end{bmatrix} \begin{bmatrix} \mathbf{A}-\mathbf{I} & \mathbf{B} \\ \mathbf{C} & \mathbf{0} \end{bmatrix} \right\} \begin{bmatrix} \mathbf{x}(k) \\ \mathbf{u}(k) \end{bmatrix} \end{aligned} \quad (10)$$

where  $\mathbf{I}$  is the matrix of ones on the main diagonal.

If we compare of the matrix of the closed loop from (7) and (10)

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ -\mathbf{KA} & -\mathbf{KB} \end{bmatrix} = \mathbf{I} + \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{K}_1 & -\mathbf{K}_2 \end{bmatrix} \begin{bmatrix} \mathbf{A}-\mathbf{I} & \mathbf{B} \\ \mathbf{C} & \mathbf{0} \end{bmatrix} \quad (11)$$

we can reach an equal dynamics (equal poles of the characteristic equation of the closed loop), if holds:

$$[-\mathbf{K}_1 \quad -\mathbf{K}_2] \begin{bmatrix} \mathbf{A}-\mathbf{I} & \mathbf{B} \\ \mathbf{C} & \mathbf{0} \end{bmatrix} = [-\mathbf{KA} \quad -\mathbf{I} - \mathbf{KB}] \quad (12)$$

From the equation (12) we can calculate the unknown vector of the state space controller  $\mathbf{K}_1$  and the gain  $\mathbf{K}_2$

$$[\mathbf{K}_1 \quad \mathbf{K}_2] = [\mathbf{KA} \quad \mathbf{I} + \mathbf{KB}] \begin{bmatrix} \mathbf{A}-\mathbf{I} & \mathbf{B} \\ \mathbf{C} & \mathbf{0} \end{bmatrix}^{-1} \quad (13)$$

#### A. Implementation of modified LQ algorithm based on the state space model to MATLAB

The modified algorithm of LQ optimal control for MIMO dynamic system is designed with using S-functions in programmable environment Matlab/Simulink. The algorithm

for the calculating of the control input value and its application to controlled system in the  $k$ -th step:

1. **start** of the simulation
2. the calculation of the gain matrix  $\mathbf{K}(k)$  in feedback loop by recursive enumeration of the Riccati equations (4) and equations (5) in the cycle, which will stop, if is executed condition  $|\mathbf{K}(k) - \mathbf{K}(k-1)| \leq \varepsilon$ , where  $\varepsilon$  is the precision of an enumeration
3. **if**  $t = t_{final\_sim}$  **then** jump to step 11
4. the loading of the state vector  $\mathbf{x}(k)$  and the controlled outputs  $\mathbf{y}(k)$  of the system and the reference trajectory  $\mathbf{r}(k)$
5. the calculating of the gain matrixes  $\mathbf{K}_1$ ,  $\mathbf{K}_2$  from the matrix  $\mathbf{K}(k)$  in the feedback closed loop of the control system by (13)
6. calculation of the control input  $\mathbf{u}(k) = \mathbf{u}_s(k) - \mathbf{K}_1 \mathbf{x}(k) + \mathbf{K}_2 \mathbf{r}(k)$
7. the calculation of the output of the summator  $\mathbf{u}_s(k+1) = \mathbf{u}_s(k) + \mathbf{K}_2 (\mathbf{r}(k) - \mathbf{y}(k))$
8. the calculation of the matrix  $\mathbf{K}(k+1)$  by (4) and (5),
9. the constraint calculated of the control input  $\mathbf{u}(k)$  by the elements on the main diagonals of the matrices  $\mathbf{Q}, \mathbf{R}$
10. the application of the control signal  $\mathbf{u}(k)$  on the system input (8), the calculation of the state vector of the system, jump to step 3
11. **end** of the simulation.

#### IV. SIMULATION OF OPTIMAL CONTROL ALGORITHM

Designed modified LQ optimal control algorithm based on state space model was applicated on simulation Ball & Plate model.

For verification of the modified LQ algorithm by designed control structure on Fig. 2 in language Matlab/Simulink was created model of the Ball & Plate.

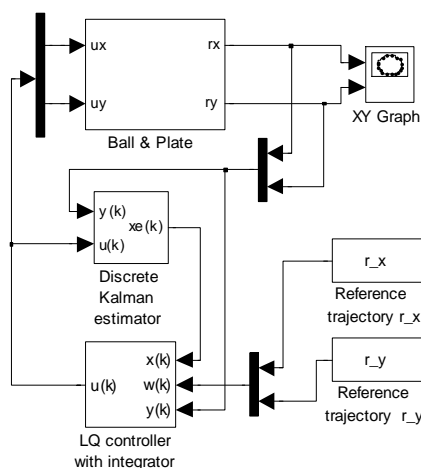


Fig. 2 The control simulation scheme

The Ball & Plate is an unstable two dimensional system with second-order astatism, which consists of a plate pivoted at its centre, such that the slope of the plate can be manipulated in two perpendicular directions. A servo system consisting of motor controller card and two stepper motors are

used for tilting the plate. Intelligent vision system is used for measurement of a ball position.

The Ball & Plate system is a dynamic system with two inputs and two outputs (Fig. 3). Both coordinates can be controlled independently as their mutual interactions are negligible due to low velocity and acceleration rate of the ball movement. The system is naturally sampled as both actuators and sensor are of a digital, discrete time nature. The system is designed to be controlled by digital controllers.

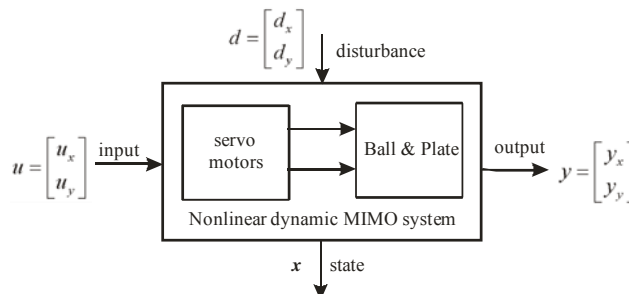


Fig. 3 The structure of Ball & Plate model – inputs/outputs

The inputs of the system Ball & Plate are the voltages of the particular motors, which are constrained between  $\langle +1, -1 \rangle$  V:

$u_x(t)$  - the voltage of the motor in the direction of  $x$  axis

$u_y(t)$  - the voltage of the motor in the direction of  $y$  axis, which ensure winding of the plate.

The other input to the system (the disturbance value) is the external force into swing out of the plate in the random direction  $d_x(t), d_y(t)$ .

The outputs of the system are actual position of the ball:

$y_x(t)$  - in the direction of the  $x$  axis,

$y_y(t)$  - in the direction of the  $y$  axis, the position of the ball is consider in the co-ordinates  $[x, y]$ .

The description of the dynamic of nonlinear lab model Ball & Plate and it's linearization form in the state space was used from manual fy Humusoft [2].

The basic control task is to control the position of a ball freely rolling on a plate. In this paper are presented the results of the tracking chosen reference trajectory (the square) by the ball using modified algorithm of the optimal control.

The parameters of the simulation at the tracking of reference trajectory have adequate selected values:

- the equilibrium point of the Ball & Plate model,
- the sapling period  $T_s$  ( $T_s = 0.15s$ ),
- the initial position of the ball in the co-ordinate system  $y_{x_{poc}} = -0.2, y_{y_{poc}} = -0.2$ ,
- the defined form of the reference trajectory  $y_{ref}(t)$
- the weight coefficients of matrices  $\mathbf{Q}$  and  $\mathbf{R}$  of the cost function  $J_{LQ}$  (3).

The results of the tracking of the reference trajectory  $r_x(t)$  and  $r_y(t)$  by the outputs of the nonlinear dynamic system

Ball & Plate  $y_x(t)$  and  $y_y(t)$  using LQ algorithm with summator are on Fig. 4 and Fig. 5. The outputs of the LQ controller – the optimal control inputs  $u_1(t)$  and  $u_2(t)$  are on Fig. 6. The tracking of the star as reference trajectory by the simulation model Ball & Plate is on Fig. 7.

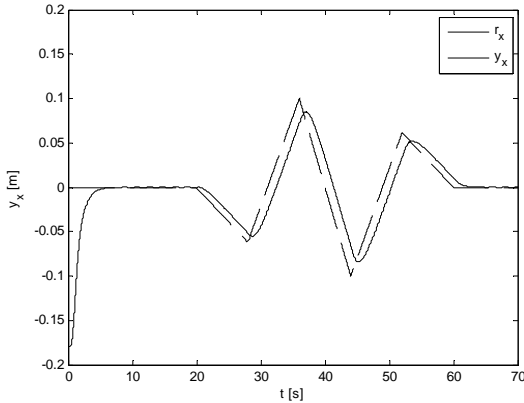


Fig. 4 The position of the ball in the direction of the  $x$  axis using modified LQ algorithm

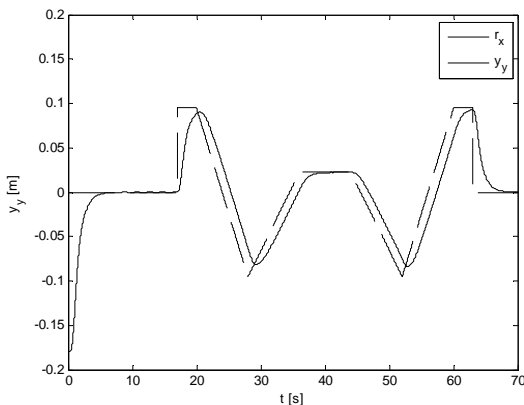


Fig. 5 The position of the ball in the direction of the  $y$  axis using modified LQ algorithm

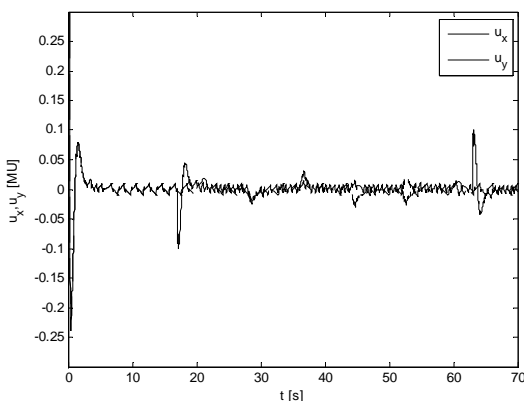


Fig. 6 The modified LQ controller outputs – control signals  $u_1(t)$ ,  $u_2(t)$

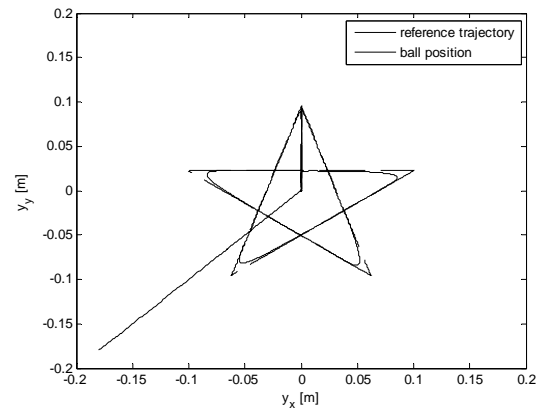


Fig. 7 The position of the ball on the plate while tracking the star trajectory

## V. CONCLUSION

The paper presents results obtained by the modified LQ optimal control algorithm design for solving problem of the tracking of the reference trajectories with application to nonlinear simulation model Ball & Plate, which are programmable using simulation language Matlab/Simulink.

## ACKNOWLEDGMENT

This research has been supported by the Scientific Grant Agency of Slovak Republic under project Vega No. 1/0617/08 *Multiagent Network Control Systems with Automatic*.

## REFERENCES

- [1] Havlena, V., Štecha, J.: „Modern Control Theory“, ČVUT Ltd., Praha, 1996, ISBN 80-01-01076-7
- [2] Humusoft: „CE151 Ball and Plate Apparatus – Educational Manual“, 1996-2004
- [3] Jadlovská, A.: „Modelling and Control Dynamic Prozesse Using Neural Networks“, Edition of the Scientific Monographs, FEI TU Košice, Informatech, Ltd., 173 pages, 2003, ISBN 80-88 941-22-9
- [4] Jadlovská, A. – Lonščák, R.: „Design and Experimental Verification of Optimal Algorithm of Control for Education Model of Mechanical System“, ElectroScope, electronic magazine for electrotechnics a electronic, No.1., Volume 2008, FEL – ZČU Plzeň 2008, p.10, Czech Rep., Internet: <http://electroscope.zcu.cz>, ISSN 1802-4564, (in Slovak)
- [5] Lonščák, R.: „Intelligent modeling and control of large scale systems“, Thesis to dissertation exam, 55 pages, Technical University of Košice, Košice, 2008. (in Slovak)
- [6] Modrlák, O.: „Principles of an analysis and the synthesis in the state space“– study materials, Technical University, Liberec, 2004
- [7] Sarnovský, J., Jadlovská, A., Kica P.: „Theory of optimal and adaptive systems“, Elfa Ltd., 171 pages, 2005, ISBN 80-8086 -020 -3, (in Slovak)

# Opinion mining as yet another approach to information extraction

<sup>1</sup>Gabriel LUKÁČ

<sup>1</sup>Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

<sup>1</sup>Gabriel.Lukac@tuke.sk

**Abstract**—In this paper a short introduction of text analyzing methods is given, generally known in the community as opinion mining. The text starts with a short historical overview of opinion mining methods and gives a formal definition of the problem. Then selected methods currently belonging to the state-of-the-art are described. At the end two approaches for profiling discussion forums participants, that are very tightly related to opinion mining, are provided. The paper ends with our design of a voting algorithm for estimating authority level of discussions participants is presented.

**Keywords**—Opinion mining, web communities, information extraction, discussions

## I. INTRODUCTION

Discussion forums and blogs represent a valuable source of knowledge and information on the web. Their open nature and constant availability offer a good opportunity to users to express their opinions about any kind of objects being shared through the Web. A massive growth of opinion-rich websites, presenting e.g. reviews of products or services, radically changed the nature of making business on the Internet. Providers of commercial services want to know *what people think* about their services and normal users ask themselves the same question to support their decision making about buying goods, watching a new film, attending a driving course, etc.

## II. HISTORY

According to the [1], the textual information published on the Web can be classified into two main categories: *facts* and *opinions*. We can treat facts as “objective statements about entities and events in the world”. On the other hand “opinions are subjective statements that reflect people’s sentiments and perceptions about entities or events”. Considering separately textual data on the Web as facts and opinions, a new opportunity to research opinion-rich subjective content is being offered.

## III. FUNDAMENTALS

When we start to talk about opinion mining, it is appropriate to define the tasks more formally and introduce a fundamental terminology. According to [1] we define the following:

**Definition 1 (object):** An *object*  $O$  is an entity which can be a product, topic, person, event, or organization. It is associated with a pair,  $O : (T, A)$ , where  $T$  is a hierarchy or taxonomy of *components (or parts)* and *sub-components* of  $O$ , and  $A$  is a set of *attributes* of  $O$ . Each component has its own set of sub-components and attributes.

**Definition 2 (opinion passage on a feature):** The *opinion passage* on a feature  $f$  of the object  $O$  evaluated in  $d$  (where  $d$  is a review document) is a group of consecutive sentences in  $d$  expressing a *positive* or *negative* opinion on  $f$ .

**Definition 3:** The *holder* of a particular opinion is a person or an organization that holds the opinion.

**Definition 4 (sentiment orientation of an opinion):** The semantic orientation of an opinion on a feature  $f$  states whether the opinion is positive, negative or neutral.

## IV. SELECTED APPROACHES TO OPINION MINING

As about 15 years have passed since the problem focused to the idea of subjectivity was introduced [2], there are several terms in the opinion mining community used interchangeably, but denoting the same thing. Along with the opinion mining definition (firstly introduced in [3]), also the phrase *sentiment analysis* co-exists in parallel, for example in work of Turney [4].

On the basis of problems addressed in the community, the main research directions are *sentiment classification* and *feature-based opinion mining*.

### A. Sentiment classification

*Sentiment classification* is a classification problem which categorizes text documents into (in most cases) three categories - *positive*, *negative* or *neutral*. This can be performed on the *whole document level* or on the *sentences level*.

1) *Document level sentiment classification:* In Turney’s work [4] a simple unsupervised learning algorithm for classifying reviews as *recommended* (thumbs up) or *not recommended* (thumbs down) is presented. The overall semantic orientation of a review is being predicted by estimating the overall semantic orientation of phrases containing adjectives or adverbs. The algorithm presented in the paper consists of three steps having a written review document on its input and producing a classification result on its output. In the first step the *part-of-speech* tagger is used to identify phrases containing adjectives and adverbs. In the second step, according to the work of Hatzivassiloglou & McKeown [5] the *semantic orientation* of each phrase is extracted. In the third step, the review is classified according to the average semantic orientation of extracted phrases.

2) *Sentence level sentiment classification:* Hatzivassiloglou and Wiebe in [6] studied language aspects that refer to expressing opinions at the level of sentences. Within the computational task they address, they distinguish sentences used to present opinions from sentences used to objectively

present factual information. In [7] Kim and Hovy study the following problem: Given a topic and a set of texts about the topic, find the sentiments expressed about the topic in each text and identify the opinion holders of each sentiment. The work of Ding et al. [1] goes deeper and tries to find opinions claimed about each feature.

### B. Feature-based opinion mining

According to Liu [8], every object can be represented by a finite set of features (e.g. frequency of CPU, energetic class of washing-machine, fuel-consumption of cars, etc.), which include the object itself. Each feature can be expressed by a finite set of words or phrases, that are synonyms. In an evaluative document, which evaluates a particular object, an opinion holder comments on a subset of the features. For each feature that opinion holder comments on, he/she chooses a word or phrase from a corresponding finite set of synonyms to describe the feature and expresses a positive, negative or neutral opinion on that feature. In the *feature-based opinion mining* the task is to discover all these hidden pieces of information from a given evaluative document.

More formally: Given an evaluative document  $d$ , the *mining result* is a set of quadruples. Each quadruple is denoted by  $(H, O, f, SO)$ , where  $H$  is opinion holder,  $O$  is object,  $f$  is a feature of the object and  $SO$  is semantic orientation of the opinion expressed on feature  $f$  in a sentence of  $d$ . Neutral opinions are ignored in the output as they are not usually useful.

The algorithm for *feature-based opinion mining* described in [8] consists of three steps. The first step is the identification of *object features*. To accomplish this task several approaches exist, for example an unsupervised method is mentioned in [9]. The second step *determines opinion orientation* on features identified in the previous step. On this task a *lexicon-based approach* using opinion words and phrases in a sentence to determine the orientation of an opinion on each feature seems to perform well. Details about algorithms can be found in [1] or [9]. Finally, the third task *grouping synonyms* is performed. It groups object features that have the same meaning but are expressed with different words or phrases.

## V. OPINION-CONDITIONAL INFORMATION EXTRACTION

Since the task of extracting opinions from textual data is strongly subjectively-oriented, there is another challenge appearing. It has become a common practice that web sites offering discussions or blog services serve as interesting media for communities to grow up around them. Thus, it makes sense to analyze the behaviour of their users and observe how their behaviour (type of contributions, semantic orientation of subjective contributions, etc.) affects how they are perceived by others - in other words how big is their *influence* or *authority* in such community?

Related to this kind of problem is the work of Matsumura et al. [10], where the *influence diffusion model* is used to discover influential comments, participants and terms from threaded on-line discussions. The model for profiling participants in on-line discussions defines *influence* as *the degree of terms propagating throughout the comment-chain of a discussion thread*. The algorithm considers that the more a comment affects other comments, the more its influence increases. Having this on mind, the influence of a subject (a comment,

participant or term) to the community is then measured as the sum of influences diffused from the subject to all other members of the community.

### A. Voting algorithm for discussion threads evaluation

The *voting algorithm* [11], that we integrated into the project SAKE [12] to analyze discussions and rank users according to their expertise level is based on an idea that discussions can be understood as an arena where social relationships emerge and take effect. The relations are formed in discussions within discussion threads. Participants obtain some (degree of) authority representing their status or position within a discussion group community. The authority of an author whose contributions are discussed depends not only on the number of people who are attracted by the contributions of this author but on the authority of these respondents as well. The authority of the author increases more if he/she is able to attract attention of more important group members than in case when his/her contributions are discussed only by group 'outsiders' - people who contribute only sparsely and their contributions do not draw attention of other group members. Initializing a new discussion thread represents a desire of an author to increase his/her authority and to strengthen the author's position within the community discussion space. Responding to a contribution of another author represents voting of a respondent for the author of the contribution and increasing the authority of the author (the contribution of the author is worth for the respondent to react).

### B. Problem representation

From the point of representation, a discussion thread is a tree-like structure nodes of which represent particular contributions while each arc represents a relationship among two contributions, one playing the role of an initiator and the second one the role of a respondent. If a contribution responds to another contribution, it is a respondent. If it attracts at least one response, then it is an initiator. It means that the root node of a tree (consisting of more than one node) representing a thread is an initiator, leaf nodes are respondents and intermediate nodes play the both roles - they are initiators and respondents at the same time. In order to process a discussion thread and to update weights of participants, a two-step iteration procedure is involved:

- Computing popularities of contributions
- Updating weights of participants

### C. Computing popularities of contributions

In order to compute popularities forming a discussion thread, a bottom-up approach is used. First, the popularity of leaf contributions (those, which do not have any respondents) is set to 1. Additionally popularities of contributions that are initiators to its respondents are computed according to the (1) influenced by the popularities of respondent contributions and weights of contributing participants.

$$y_j = \sum_{i=1}^k w_i y_i \quad (1)$$

The popularity of the root contribution  $y_j$  is computed as a sum of popularities  $y_i$  of  $k$  respondents whose authors have quantitative authority weights  $w_i$ . Using this recursive

approach the popularity of leaf contributions is computed at first. The popularity of the root initiator contribution is determined at the end of the cycle.

#### D. Updating weights of participants

The second part of the algorithm is the authority weight update of participating authors in a given discussion thread. The weight modification depends on how many contributions authors have authored, how popular their contributions are, as well as how many contributors were attracted by the given thread. The weight of an author is updated for each of his/her contributions in the thread according to the formula (2),

$$w_k^{n+1} = w_k^n + \frac{y_i}{\bar{y}} \frac{n}{N} \quad (2)$$

where  $w_k$  is the authority weight of author  $k$ ,  $\bar{y}$  is the average popularity of contributions in a given thread,  $n$  is the number of participating authors in the thread and  $N$  is the size of the community.

When analyzing discussions we deal with a wide portfolio of opinions, where expressing consensus or critique could be perceived as moving factors for suggesting new themes, trends, etc. Applying real-world conditions to our additive voting model it seems to be logical, that not every response contribution represents a positive vote to a parent one. Applying this way of thinking, dealing with critique contributions should be also incorporated into the voting model for computing authorities of participants in order to make it more realistic.

Keeping these problems in our mind, we have extended the voting approach to discussion threads evaluation algorithm with a third step [13], so the final form of the algorithm is as follows:

- Calculating popularity of contributions for a given discussion thread
- Modification of contributions popularities based on received user feedback values
- Updating weights of contributors

The user feedback is collected for every message manually in the manner, that every discussion forum user poses a possibility of one-time evaluation of any message by selecting a value from a list of predefined choices (0 - Bad, 1 - Rather Bad, 2 - Neutral, 3 - Rather Good, 4 - Good) that represent his/her opinion about the concrete message. On the basis of collected user feedbacks for particular messages, new message popularity values are computed:

$$y'_j = 0,25\bar{f}_j y_j \quad (3)$$

where  $y'_j$  is a new popularity value of the contribution  $j$ ,  $\bar{f}_j$  is an average value of user feedbacks for the contribution  $j$ ,  $y_j$  is a previous popularity value of the contribution  $j$  calculated not considering users' feedback. By using this new proposed approach, the original popularity value  $y_j$  of message  $j$  is transformed to a new value  $y'_j$  from interval  $\langle 0, y_j \rangle$  according to the enriched user feedback value of the original post. If there is no user feedback value for contribution  $j$ , the average value  $\bar{f}_j$  is set to a neutral value of 2 and this new setting is used when the new popularity is being computed. In case of omitting this step all the contributions without any user feedback are perceived as very competent contributions, but it might not be true.

## VI. CONCLUSION

In the paper several methods that represent state-of-the-art in the field of opinion mining were presented. From our point of view especially blogging communities can be seen as a valuable source of opinion-rich data providing opportunities for extracting information about profiles of influential authors. We also presented a voting algorithm for estimating authority-level of threaded discussions participants. As a further work on this algorithm we propose to extend the current version collecting the user-feedback manually by an automatic opinion mining extension, which will extract the feedback automatically from contributions in a comment-chain.

## ACKNOWLEDGMENT

The work presented in the paper was supported by the Slovak Grant Agency of the Ministry of Education and Academy of Science of the Slovak Republic within the 1/4074/07 Project "Methods for annotation, search, creation, and accessing knowledge employing meta-data for semantic description of knowledge".

## REFERENCES

- [1] X. Ding, B. Liu, and P. S. Yu, "A holistic lexicon-based approach to opinion mining," in *WSDM '08: Proceedings of the international conference on Web search and web data mining*. New York, NY, USA: ACM, 2008, pp. 231–240.
- [2] J. M. Wiebe, "Tracking point of view in narrative," *Computational Linguistics*, vol. 20, pp. 233–287, 1994.
- [3] K. Dave, S. Lawrence, and D. M. Pennock, "Mining the peanut gallery: opinion extraction and semantic classification of product reviews," in *WWW '03: Proceedings of the 12th international conference on World Wide Web*. New York, NY, USA: ACM, 2003, pp. 519–528.
- [4] P. D. Turney, "Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews," in *ACL '02: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*. Morristown, NJ, USA: Association for Computational Linguistics, 2001, pp. 417–424.
- [5] V. Hatzivassiloglou and K. R. McKeown, "Predicting the semantic orientation of adjectives," in *Proceedings of the eighth conference on European chapter of the Association for Computational Linguistics*. Morristown, NJ, USA: Association for Computational Linguistics, 1997, pp. 174–181.
- [6] V. Hatzivassiloglou and J. Wiebe, "Effects of adjective orientation and gradability on sentence subjectivity," 2000. [Online]. Available: [citeseer.ist.psu.edu/hatzivassiloglou00effects.html](http://citeseer.ist.psu.edu/hatzivassiloglou00effects.html)
- [7] S.-M. Kim and E. Hovy, "Determining the sentiment of opinions," in *COLING '04: Proceedings of the 20th international conference on Computational Linguistics*. Morristown, NJ, USA: Association for Computational Linguistics, 2004, p. 1367.
- [8] B. Liu, "Opinion mining," *Invited contribution to Encyclopedia of Database Systems, to complete in July*, 2008.
- [9] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *KDD '04: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. New York, NY, USA: ACM, 2004, pp. 168–177.
- [10] N. Matsumura, Y. Ohsawa, and M. Ishizuka, "Profiling participants in online-community based on influence diffusion model," *Transactions of the Japanese Society for Artificial Intelligence*, vol. 18, pp. 165–172, 2002.
- [11] M. Mach and G. Lukáč, "A dedicated information collection as an interface to newsgroup discussions," in *IIS 2007 : 18th international conference on Information and Intelligent Systems*. Faculty of Organization and Informatics, University of Zagreb, Sept. 2007, pp. 163–169.
- [12] G. Butka, Peter Lukáč, "Architecture and integration of prototype in sake," in *Znalosti 2008 : 7. ročník konferencie*. Bratislava: STU, 2008, pp. 327–330.
- [13] P. Butka and G. Lukáč, "Semantic-based groupware system in sake for support of knowledge and expert intensive public administration processes," in *In: CECIIS 2008 : Proceedings of the 19th Central European Conference on Information and Intelligent Systems*. University of Zagreb, 2008, pp. 307–314.

# Design of dataflow computer architecture with tile organization

<sup>1</sup>*Branislav MADOŠ*

<sup>1</sup>Dept. of Computer and Informatics, FEI TU of Košice, Slovak Republic

Branislav.Mados@tuke.sk

**Abstract**— This article deals with tile computing as a modern approach to multi-core design of microprocessors and describes an effort to design architecture of computer with dataflow computing paradigm in combination with tile computing paradigm. This work comes out from the research in the field of dataflow computing and architectures of dataflow computers DF KPI and DFC1 at the Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics, Technical University of Košice. This work is supported by VEGA Grant No. 1/4071/07 and APVV-0073-07.

**Keywords**— Dataflow computing, Tile computing, Spatial Computing.

## I. INTRODUCTION

Exponentially increasing amount of transistors integrated on the chip with increasing density of integration, in accordance to the Moore's law, provides opportunity to increase the performance of microprocessors.

Conventional superscalar architecture of microprocessor cannot be simply scaled up in line with this trend, and it results in challenging problem to maintain continual proportion between amount of integrated transistors and performance of the chip.

Disproportion between communication performance and computation performance intensifies with miniaturization of transistors, because of side effect of relative lengthening of inside chip wiring. Centralized control of microprocessor intensifies the problem. Memory model using cache memory causes extensive associative searches.

Increasing circuit complexity of superscalar microprocessors with longer design times, complex tests of designs and increasing amount of design bugs are also negative aspects of effort to increase the performance of superscalar microprocessors by integration of more transistors into single processor core.

Mainstream multi-core designs of superscalar processors are trying to address those challenges and commercially available microprocessors are integrating two or four cores into single chip.

Pushing trend in the architecture of multi-core microprocessors is integration of tens of cores in tiled configuration and this approach is called tile computing paradigm.

## II. TILE COMPUTING

Representatives of general purpose tile computing microprocessors are Tile64[1] designed by Tiler Corporation, integrating 64 general purpose cores called tiles. Each core integrates L1 and L2 cache. Tiles are arranged in 8 x 8 bidimensional mesh using interconnecting network with 31Tbps data throughput. Chip utilizes 3-way VLIW pipeline for instruction level parallelism. Each tile can independently run operating system, or multiple tiles can run multiprocessing operating system. Performance of the chip at 700 MHz is  $443.10^9$  Operations Per Second (BOPS).

Intel microprocessor TeraScale Processor is designed under Tera-Scale Computing Research Program[2]. Terascale Processor integrates 80 cores in bidimensional mesh in 8 x 10 cores organization. With 65 nm processor technology implemented, chip integrates 100 million transistors on 275 mm<sup>2</sup> die. Processor performs 4 FLOP per cycle per core, and with 4.27 GHz delivers theoretical peak performance of 1.37 TFLOPS in single precision. Instructions set architecture consists of 16 instructions and Terascale Processor uses 96 bit VLIW. Each core runs own program and interconnection network transfers data and coordination instructions for execution of programs between cores via message passing.

Representative of DSP microprocessors with tile organization is VGI described in [3]. Organization of 64 cores is 8 x 8 in bidimensional mesh and each core contains approximately 30 000 transistors. VGI utilizes dataflow paradigm of computation.

Whereas described architectures are integrating tens of cores, new approach to microprocessor design represented by spatial computing, with TRIPS, RAW, SmartMemories, nanoFabrics or WaveScalar representatives, integrates hundreds or even thousands of simple cores or processing elements (PE) on the chip, often arranged along with memory elements in the grid [3][4].

### III. DATAFLOW COMPUTING

Dataflow paradigm of computation was popularized in 60' and 70' by [5][6][7] and is non Von Neumann architecture with the ability of fine grain parallelism in computation process.

In dataflow architecture the flow of computation is not instructions flood driven. There is no concept of program counter implemented in this architecture. Control of computation is realized by data flood. Instruction is executed immediately in condition there are all operands of this instruction present. When executed, instruction produces output operands, which are input operands for other instructions.

Dataflow paradigm of computing is using directed graph  $G = (V, E)$ , called DataFlow Graph (DFG). DFG is used for the description of behavior of data driven computer. Vertex  $v \in V$  is an actor, a directed edge  $e \in E$  describes precedence relationships of source actor to target actor and is guarantee of proper execution of the dataflow program. This assures proper order of instructions execution with contemporaneous parallel execution of instructions. Tokens are used to indicate presence of data in DFG.

Actor in dataflow program can be executed only in case there is a presence of a requisite number of data values (tokens) on input edges of an actor. When firing an actor execution, the defined number of tokens from input edges is consumed and defined number of tokens is produced to the output edges.

An important characteristic of dataflow program is its ability to detect parallelism of computation. This detection is allowed on the lowermost basis – on the machine instructions level.

There are static, dynamic and also hybrid dataflow computing models.

In static model, there is possibility to place only one token on the edge at the same time. When firing an actor, no token is allowed on the output edge of an actor. Disadvantage of the static model is impossibility to use dynamic forms of parallelism, such a loops and recursive parallelism. Computer with static dataflow architecture was designed by Denis and Misunas and introduced in [5].

Dynamic model of dataflow computer architecture allows placing of more than one token on the edge at the same time. To allow implementation of this feature of the architecture, the tagging of tokens was established. Each token is tagged and the tag identifies conceptual position of token in the token flood.

For firing an actor execution, a condition must be fulfilled that on each input edge of an actor the token with the same tag must be identified. After firing of an actor, those tokens are consumed and predefined amount of tokens is produced to the output edges of an actor. There is no condition for firing an actor that no tokens must be placed on output edge of an actor. The architecture of dynamic dataflow computer was first introduced at Massachusetts Institute of technology (MIT) as a Tagged Token Dataflow Architecture in [6].

Hybrid dataflow architecture is a combination of control flow and data flow computation control mechanisms.

Dataflow computing is predominantly domain of the research laboratories and scientific institutions, and has

limited impact on commercial computing because of difficulties in cost of communication, organization of computation, manipulation with structured data and cost of matching.

Paradigm of tile computing in combination with dataflow computing brings new possibilities to overcome some of deficiencies of dataflow architectures.

### IV. PROPOSED ARCHITECTURE

Proposed architecture consists of elements with simple design, represented by processing elements (PE), Input/Output elements (I/O), and two interconnection networks spread across the chip.

Processing elements are arranged in accordance with tile computing paradigm into bidimensional mesh across the whole chip. Each PE integrates registers in function of data and instruction code stores, arithmetic and logic unit for instruction execution and control unit. All PEs are unified general purpose computing units with very simple design.

Input and Output (I/O) elements connected to the pins of the chip are localized at edges of the mesh of processing elements.

First of interconnection networks allows short range high-bandwidth communication between neighbor PEs. In this communication network processing element communicates directly with eight neighbors and allows speed communication over short distance. Processing elements at edges of the PE mesh are able to communicate with I/O elements. Short range communication network allows concurrent communication of elements of the chip.

Second interconnection network is spread along the chip allowing long range communication which interconnects all PEs to all PEs and I/O elements of the chip. Communication on the long range communication network is limited to two elements of the chip; other components of the chip cannot communicate through the long range communication network at the same time.

PE is able concurrently communicate over short distance and long distance network, with limit to 2 inputs and 2 outputs concurrently.

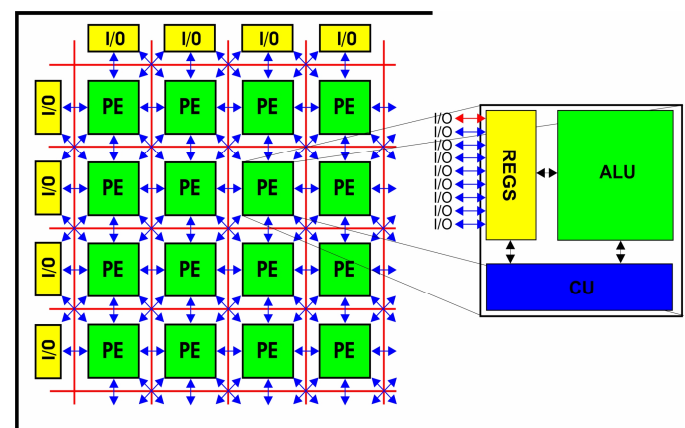


Fig. 1. Organization of the microarchitecture of proposed dataflow computer, and processing element.



Special interface of the chip allows mapping dataflow graph onto the chip elements, which registers are addressable as a memory cells and whole chip comes across as a memory chip.

Load of dataflow graph (DFG) is realized through the long distance communication network and DFG is mapped onto the elements of the chip. Each instruction is mapped onto one of PEs and information on source and target elements mapped into registers of elements establishes communication lines for data flow.

In geometrical point of view, mapping of DFG means assigning of vertexes of DFG to the PEs and establishing edges by definition of source and target elements addresses.

It is impractical to map all DFG into the chip and that is why it is possible to dynamically map instructions into PEs. Instructions and communication lines are dynamically replaced when unneeded by newly activated. Crucial for effective mapping of DFG is to map instructions in the manner in which producer and consumer of token are directly interconnected via short range communication network with aim to establish interconnections with small latency.

Advantages of architecture with tile organization are in simple design of elements which are arranged across the chip in uniform simple manner which secures high-level scalability of the architecture. Decentralization of data storing, execution and control of computation results in shorter wire lengths on the chip and small latency in communication. Tile based organization is able to tolerate some defects in manufacturing of the chip, because it is possible to detach erroneous tiles from the mesh.

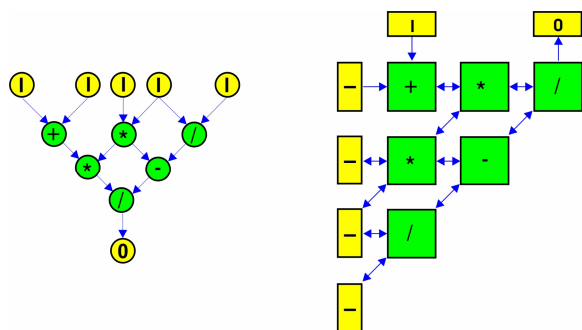


Fig. 2. Mapping of the dataflow graph onto the computer structures.

## V. TECHNOLOGY

Prototype of dataflow computer with tile based architecture is realized as the software simulation with use of development software tool ISE WebPack for architecture development and a software tool ModelSim for simulation and verification of function of realized design. Simulation is oriented to Spartan 3 FPGA chip with the aim to prepare hardware prototype of the computer with use of Xilinx Spartan 3 PCIe Starter board.

There is an FPGA chip Xilinx Spartan 3 XC3S1000-4FG676 with 676 pins in FBGA package in the center of

development board with 50 MHz clock frequency. There are 391 free for use pins in this chip and 17000 logic cells.

Development board involves 4 Mb serial flash memory. There are 8 light emitting diodes (LED) for diagnostics and development purpose on the design board. There is also 9-pin RS 232 serial interface and 168 pins expansion ports which functions are configurable in accordance with the design on the board.

Development tools were selected with consideration of the availability and above all rate and simplicity of the iterative process of design, implementation and verification cycle.

## VI. CONCLUSION

Contribution of this work is in the design of the architecture and realization of software simulation of dataflow computer with the tile based architecture, which allows further research in the field oriented on possibilities of utilization of dataflow computing in praxis and brings possibilities to overcome some disadvantages of dataflow paradigm. This work introduced paradigm of tile based computing, dataflow computing and introduced details of dataflow computer architecture designed at the Department of Computers and Informatics, Technical University of Košice, with support of VEGA Grant No. 1/4071/07 and APVV-0073-07.

This work was supported by VEGA Grant No. 1/4071/07 and APVV-0073-07.

## REFERENCES

- [1] Dawid Wintzlaw, Patrick Griffin, Henry Hoffmann, Liewei Bao, Bruce Edwards, Carl Ramey, Matthew Mattina, Chyi-Chang Miao, John F. Brown III, Anant Agarwal, "On-Chip Interconnection Architecture of the Tile Processor," IEEE Micro, vol. 27, no. 5, pp. 15-31, Sep./Oct. 2007, doi:10.1109/MM.2007.89.
- [2] Mattson, T. G., Wijngaart, R., Frumkin, M. „Programming the Intel 80-core network-on-a-chip terascale processor“, Conference on High Performance Networking and Computing, Proceedings of the 2008 ACM/IEEE conference on Supercomputing, 2008, pp. ISBN 978-1-4244-2835-9.
- [3] Mercaldi, M., Swanson, S., Petersen, A., Putnam, A., Schwerin, A., Oskin, M., Eggers, S., J. "Modeling Instruction Placement on a Spatial Architecture", SPAA'06, July 30–August 2, 2006, Cambridge, Massachusetts, USA, 1595934529/06/0007.
- [4] Swanson, S., Schwerin, A., Mercaldi, M., Petersen, A., Putnam, A., Michelson, K., Oskin, M., Eggers, S., „The WaveScalar Architecture“.
- [5] J. B. Dennis and R. P. Misunas, "A Preliminary Architecture for a Basic Data Flow Processor", Proceedings of the 2nd Annual Symposium on Computer architectures., 1974.
- [6] J. Sharp, "Data Flow Computing", Ellis Horwood, West Sussex, England, 1985.
- [7] A. Veen, "Dataflow Machine Architecture", ACM Computing Surveys, December 1986, pp. 365-396.
- [8] L. Vokorokos and N. Ádám, "Operators Matching in Dynamic Data Flow Architectures", Conference on Computation Cybernetics, Vienna, Austria, August 30 - September 1, Vienna, 2004, pp. 77-81, ISBN 3-902463-02-3.
- [9] L. Vokorokos, N. Ádám and A. Baláz, "Flexible Platform for Neural Network Based on Data Flow Principles", HUCI, Budapest, November 18-19, Budapest Tech, 2005.
- [10] L. Vokorokos, "Parallel Computer System Utilization in Geographic Information Systems", IEEE 3rd International Conference on Computational Cybernetics, Maurícius, April 13 - 16, Budapest, 2005, pp. 333-338, ISBN 963 7154 37 X.

# Mobile content delivery using Videoservert

Miroslav MICHALKO

Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

miroslav.michalko@tuke.sk

**Abstract** — This paper deals with implementation of latest streaming technology developments into experimental testbed of Computer Networks Laboratory at Technical University of Kosice. Paper describes in brief current state of art and presents available of mobile streaming solutions for students. User interaction is offered through Videoservert platform which provides services for both, desktop and mobile users. Several techniques of QoS and cross-platform interoperability were implemented in order to improve services to end user. Moreover this platform combines on-demand streaming capabilities with modern videoconferencing solutions. This paper is dedicated to ten years anniversary of work on this platform, seven years by author.

**Keywords**—streaming, mobile, content delivery, iphone.

## I. INTRODUCTION

Recent development in computer networks and mobile technologies opened wide field of opportunities for delivery of high quality multimedia content for many users with end devices which are capable of high processing power and high speed connection to network. With these underlying technology capabilities there is opened field for streaming as universal content delivery method. Streaming is a method of video/audio delivery in real-time in a way that the recipient can view transferred content while receiving rest of the video/audio file. Basically streaming is not limited only to audio or video, but this paper will focus only on this content. Streaming allows delivery of any content. Advantages of this type of delivery are clear. End users do not need to wait for whole time period to download and store multimedia file. But on the other side, quality and smoothness is given by network infrastructure. Video quality defined by its size, bit rate and other parameters needs to convert to end device and network bandwidth limitations.

Videoserver was developed several years ago as simple streaming service to end users offering unicast streaming using Windows Media Services as streaming core technology. With rapid advancements of computers and networks there was need to implement better solutions which allows wider audience to access multimedia content. Platform provides powerful and easy-accessible services for sharing video and audio resources using Flash technology. For mobile clients there are several possibilities how to deliver content. Content is streamed in optimized data stream. Videoserver allows also sharing of audio files and pictures. One of powerful service provided by Videoserver is possibility to create multimedia content in supporting environment. After accessing archive,

Videoserver initiates streaming of video stored in server directly to your computer over the network. Videoserver also contains several technologies which support access to formats using mobile devices.

## II. VIDEOSERVER - MOBILE STREAMING

With the development of broadband wireless networks, this makes it feasible to have mobile streaming receivers and mobile streaming sources. In fact, third-generation cellular networks are designed to allow mobile devices to move while receiving IP traffic. The range of mobile devices that already use streaming media is growing and will continue to grow. Videoserver now offers streaming to wide variety of mobile phones. As core streaming technology Darwin Streaming Server is used (DSS). DSS uses H.264 container for delivering of content and is platform independent what makes it excellent tool for streaming in high quality to all streaming enabled mobile devices [1].

Quite hard task is to implement mobile streaming to Apple iPhone. The iPhone is modern device with high computing power but unfortunately with partially closed documentation related to streaming media. Thanks to some open source web projects and reverse techniques a way was found and Videoserver now offers specially built user friendly web application to access archive. This is second stage in plan to allow streaming of educational content to everyone everywhere. The iPhone website is dynamically generated and all changes in video archive are mirrored at the same time. Voting, commenting and other features of full Videoserver version are going to be added in later version of iPhone website.

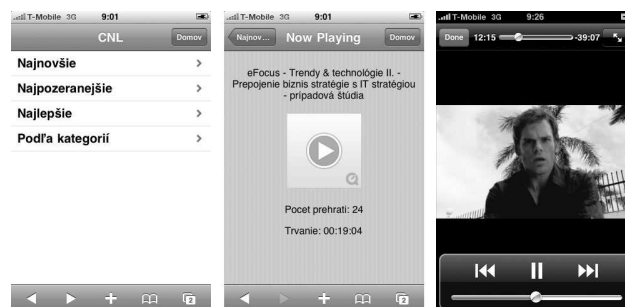


Fig. 1. Videoserver – iPhone user interface

Videoserver also offers automatic end device identification based on User Agent field which is part of HTML header. It is critical to precisely determine type of end client in order to

provide specially optimized data stream. The iPhone has fully different method of accessing content. Even if it has fully compatible web browser, touch screen is simpler to use with dedicated larger scale website. As streaming method only HTTP streaming is used. DSS is for this type of phone bypassed for incompatibility issues. For all other types of mobile devices DSS is used. The iPhone also need separate video format MP4 so duplicate video content is stored by Videosever platform. Videosever stores content primary in Flash format and unfortunately iPhone plays this format only in YouTube application. For this reason MP4 is only playable format supported in external application, in this case it is web browser [2].



Fig. 2. Videosever – Desktop user interface (non-mobile access)

### III. VIDEOSERVER – AUTOMATIC CONTENT PROCESSING

Videosever allows automatic video processing using FFmpeg. FFmpeg is a collection of software libraries that can record, convert and stream digital audio and video in numerous formats. It includes libavcodec, an audio/video codec library used by several other projects, and libavformat, an audio/video container mux and demux library. After submitting video using portal Upload form ffmpeg is used to automatically convert video into flash format. During converting phase user is allowed to see partial video output of encoded video [3].

Videosever streams data flows using Flash coded video and HTTP protocol, most simple and cheapest method. HTTP streaming is enough for our purposes, but for larger video portals it's recommended to use serious streaming technology.

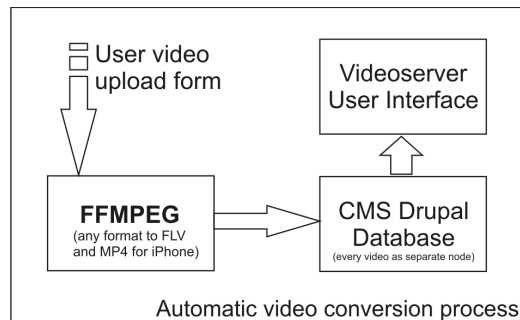


Fig. 3. Videosever - Automatic video processing of user video content

As client JW FLV is used and for streaming purposes Lighttpd server instead on Apache. In case of higher number of clients Red5 technology will override HTTP streaming. Red5 is a free, open source Flash server that supports streaming and recording of audio/video, live stream publishing and Flash remoting. Big advantage of this Java based application is cooperation with FFMPEG software package [4].

### IV. VIDEOCONFERENCE AND VIDEOSERVER

As part of improvement of Technical University of Kosice infrastructure is implementation of videoconference kits into fifteen large lecture rooms. Videoconference kits supports streaming of video and content using various formats. As most used and supported is H.323. Due to better quality parameters actually preferred formats are H.264 and H.239 with content in one stream capabilities. Output stream of every videoconference kit is mapped to Videosever as video source and redistributed to end users. This functionality is only in testing and not publically accessible. Full official launch is planned till the end of this year.

Still there are some unsolved problems. Videoconference kits do not supports automatic recording of broadcasted sessions. Implementation of this functionality into Videosever is time consuming and required dedicated person to make post processing and publishing into pre-selected categories. As one of the considerable solution is to buy commercial software, for example Codian server [5]. Disadvantage of this access is limitation of proprietary solutions and interoperability with other systems. In this case problems should be with on the fly recoding to Flash stream. Flash is preferred technology in Videosever platform.

### V. HDTV STREAMING

There are only few streaming technologies available with HDTV streaming capability. Most known are Microsoft Media Services with its own proprietary HDWMV codec, Helix based solution provided by Real, QuickTime and VLC. Two most suitable technologies VLC and Helix are implemented and provide multicast and unicast on-demand streaming. For unicast streaming special module was developed and implemented into IP streaming platform running on CMS Drupal [6].

Due to limitations of multicast streaming content is accessible only within local intranet of Technical University in

Kosice - TUNET. In order not to overload network with unnecessary data stream platform allows also editing TTL value in TCP packet header. After reaching value equal to zero packet is destroyed by any active layer 3 packet device.

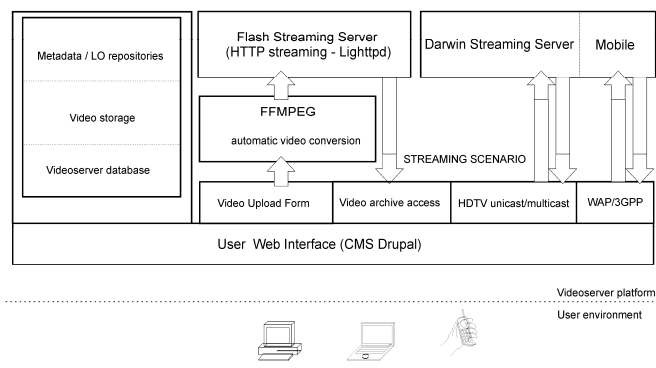


Fig. 4. Videoserver – Internal architecture overview

HDTV video streams are propagated using SAP playlist what makes system more user friendly in compare to use only IP multicast address to remember. For extranet viewers interested in multimedia content there is only on-demand streaming available using specially developed web interface. Due to higher data bandwidth requirements only users with capable connection are ready to access content [7].

## VI. FUTURE PLANS

Videoserver platform provides complex service for sharing multimedia content using latest streaming technologies. Implementation of mobile streaming extension is unique in compare to similar web portals. Next development will focus on higher level of user interactivity, streaming adaptability based on continuously changing conditions within providers' network environment, distributed access to video content and automatic metadata recognition.

Goal of this work is to make from Videoserver complex IPTV solution which will offer not only streaming of stored content but also live broadcasting. There are many challenges to be solved and many technology concepts to be tested. One of interesting work to be done is to cover live broadcasting from various large lecture rooms at university and offer to students who are not able to come to school. Also system of automatic archivation and meta tagging need to be created. This work will be part of next planned bachelor and diploma works.

## VII. CONCLUSION

Videoserver is powerful platform for sharing multimedia content to end user. As academic project it offers mostly educational content created at university, conferences and workshops. Next level of development is focused on new desktop and mobile services, interactive content and simple communication user interface but with latest streaming capabilities offering services anytime and everywhere. Technologies presented in this paper show only few examples of utilization of new streaming and processing technologies connected together in order to provide services with added

value. Videoserver is up and ready solution for educational institution for delivering content to online learners.

## ACKNOWLEDGMENT

Author would like to thank to all members of video team at Computer Networks Laboratory, mostly students working on this project for many years as part of their semestral, bachelor and diploma work.

## REFERENCES

- [1] Apple Inc., "Open Source Streaming Server," 2008, [Online; accessed 24-March-2009]. [Online]. Available: <http://developer.apple.com/opensource/server/streaming/index.html>
- [2] Wikipedia, "MP4," 2009, [Online; accessed 24-March-2009]. [Online]. Available: <http://sk.wikipedia.org/wiki/MP4>
- [3] Wikipedia, "Ffmpeg," 2009, [Online; accessed 24-March-2009]. [Online]. Available: <http://en.wikipedia.org/wiki/Ffmpeg>
- [4] Red5, "Open source flash server," 2007, [Online; accessed 25-March-2009]. [Online]. Available: <http://osflash.org/red5>
- [5] Tandberg, "IPVCR," 2009, [Online; accessed 25-March-2009]. [Online]. Available: <http://www.tandberg.com/products/ipvcr.jsp>
- [6] M. Michalko, V. Sidimak: Streaming of multimedia to mobile and desktop users using Videoserver platform, SCYR 2008, 8th Scientific Conference of Young Researchers, Košice, May 28, 2008, Košice, Faculty of Electrical Engineering and Informatics, Technical University of Košice, 2008, I. Edition, pp. 105-107, ISBN 978-80-553-0036-8
- [7] F. Jakab, M. Michalko, M. Binas, and J. Baca, "Flash and Mobile Solutions for Video Streaming Platform," 5th International Conference on Emerging e-Learning Technologies and Applications, Conference Proceedings, pp. 163–170, Sept. 2007.

# Magnetic Aura of Small Turbojet Engine MPM 20

Jana Modrovičová

Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

[jana.modrovicova@tuke.sk](mailto:jana.modrovicova@tuke.sk)

**Abstract**—This article focuses on confirmation of existence of magnetic aura of small turbojet engine MPM 20. Also the influence of MPM 20 engine to the Earth's magnetic field is described and shows how the engine force deforms Earth's magnetic field in its vicinity.

**Keywords**— magnetic aura, MPM 20, measurement, Earth magnetic field.

## I. INTRODUCTION

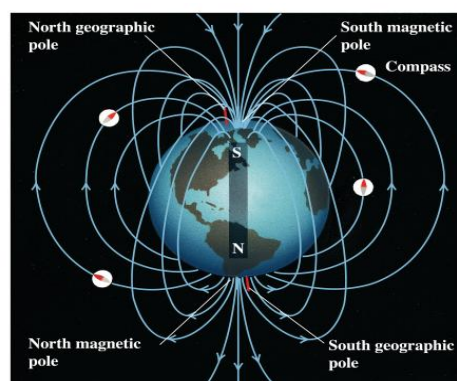
Magnetism as well as other physical attributes belongs to the most important characteristics of particles in universe. It is also important for planets and all objects in space and therefore also for the observed object. This object is small turbojet engine stationed on Faculty of Aeronautics of Technical University in Košice. More detailed description of MPM 20 can be found in works [1, 5].

Each object has its energy which is emitted to the environment and it is the same with the magnetism of MPM 20. Therefore it can be called as magnetic aura of MPM 20. Basically, the magnetic aura of MPM 20 can be interpreted as magnetic field in vicinity of the engine, where this field is characteristic for the MPM 20. At the same time can be said, that this magnetic field in neighborhood of the engine is actually Earth's magnetic field influenced by engine, or generally any object.

This article is therefore focused on examination of Earth's magnetic field in surroundings of MPM 20, hence on its magnetic aura and degree of deformation caused to the Earth's magnetic field.

## II. EARTH MAGNETIC FIELD AND EARTH'S MAGNETIC FIELD IN VICINITY OF MPM 20

Magnetic field of the Earth behaves as huge magnetic dipole and at the same time its magnetic poles are not identical with geographical poles of the Earth [8]. Next figure depicts lines of force of magnetic field, magnetic and geographical poles.



Copyright © Addison Wesley Longman, Inc.

Fig. 1. Lines of force of Earth's magnetic field, magnetic and geographical poles [10]

Next figure shows constituent components of Earth's magnetic field.

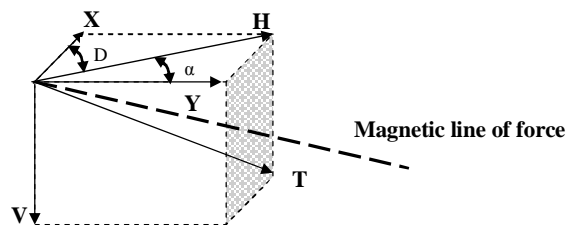


Fig. 2. Components of Earth's magnetic field [4, 8]

As can be seen on the picture, vector  $\vec{T}$  forms with line of force an angle  $\alpha$ , described as *geomagnetic inclination*. Vector of magnetic induction  $\vec{T}$  can be decomposed to horizontal  $\vec{H}$  and vertical  $\vec{V}$  component. Horizontal component isn't identical to the geographical meridian and the angle between vector  $\vec{H}$  and geographical meridian  $\vec{X}$  in certain place is called *geomagnetic declination*. Component  $\vec{Y}$  is perpendicular to geographical meridian  $\vec{X}$ . Relations between constituent components can be mathematically described by following:

$$\vec{X} = \vec{H} \cdot \cos \vec{D} \quad (1)$$

$$\vec{Y} = \vec{H} \cdot \sin \vec{D} \quad (2)$$

$$\vec{V} = \vec{H} \cdot \operatorname{tg} a = \vec{T} \cdot \sin a \quad (3)$$

$$\vec{T}^2 = \vec{H}^2 + \vec{V}^2 \quad (4)$$

$$\vec{H} = \vec{T} \cdot \cos a \quad (5)$$

$$\frac{\vec{X}}{\vec{Y}} = \operatorname{tg} \vec{D} \quad (6)$$

As was mentioned before, magnetic field is in vicinity of all objects and therefore also in neighborhood of MPM 20. That's why examination of magnetic aura can be imagined as monitoring of changes in intensity of Earths magnetic field in neighborhood of MPM 20. Magnetic field is not only time variable, but depends also strongly on magnetic properties of observed (monitored) objects. Following figure depicts observed object and chosen grid system for measurements

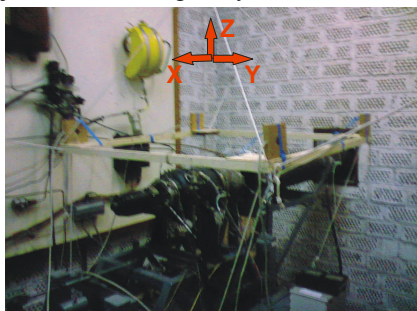


Fig. 3. Laboratory with observed object of measurement MPM 20 and chosen grid system [8]

Figure shows clearly, that magnetic field is influenced not only by examined object, but also by measurement systems. Therefore it was necessary to measure the deviation, or deformation of Earths magnetic field. Measurement was done in two phases:

**A. Measurement of outside deviation**

This measurement was done outside the building (open space), where the observed object is stationed, and chosen measuring device (compass) was aligned parallel with axis y of the chosen grid system (fig. 4).

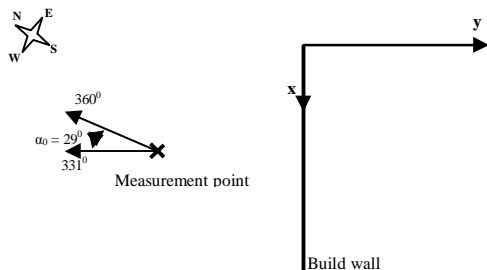


Fig. 4. Measurement of deviation of Earths magnetic field outside the building [8]

**B. Measurement of deviation in MPM 20 surroundings**

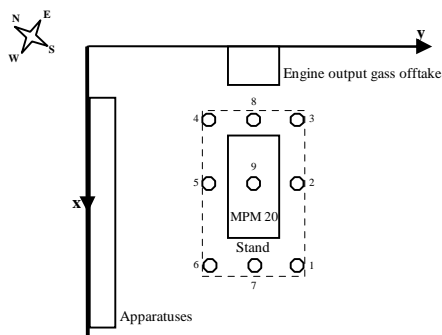


Fig. 5. Measurement of deviation of Earths magnetic field in neighborhood of MPM 20 [8]

This measurement was done in the room where the

observed object MPM 20 is stationed and the deviation was measured in nine different places around the MPM 20 and above it (fig. 5).

TABLE I  
TABLE WITH MEASURED DEVIATIONS FROM EARTHS MAGNETIC FIELD IN VICINITY OF MPM 20 [8]

Position no.	Value of deviation $\alpha_i$ [°] (i=1,2,...,9)	Value of $\alpha_0$ [°]	Difference $ \alpha_i - \alpha_0 $
1	35	29	6
2	29	29	0
3	25	29	4
4	41	29	12
5	38	29	9
6	44	29	15
7	39	29	10
8	24	29	5
9	29	29	0

Legend:

$\alpha_0$  – deviation from the Earths magnetic field outside the building

1, 2, ... 9 – locations in which the deviation was measured

x, y – chosen grid system around MPM 20

Differences between values measured in two phases are stated in the table. So it is possible to evaluate considerable deformation of Earths magnetic field in surroundings of MPM 20. To confirm the force deformation of Earths magnetic field several tests were done and their results are described in next chapter.

III. MEASUREMENTS

Two types of magnetometers were used to measure the magnetic aura of MPM 20, specifically single-component magnetometer VEMA-30 and N-channel analyzer of magnetic fields. More detailed description of mentioned magnetometers can be found in [2, 3, 6].

To confirm the existence of magnetic aura of MPM 20, several measurements were done with introduced magnetometers. In the beginning, single-component measurements were done, followed by 4-component (4-channel) measurements of so-called "night magnetic peace" and then short-term measurements. Exact measurement procedures as well as results can be found in work [8].

This article depicts results of short-term measurements done by 4-channel magnetometer and in chosen coordinate systems in 5 minutes intervals [8].

**A. Measurement in the direction of axis x**

During this measurement, probes were positioned in parallel with x axis in chosen coordinate system

In following table, meteorological conditions for this series

TABLE II  
METEOROLOGICAL CONDITIONS FOR 21.10.2008

Date	Dew point [°C]	Atmosph. pressure [hPa]	Medium external temp. [°C]	Medium internal temp. [°C]	Ozone [Dobs. unit]
21.10.	5	1022	7	8	257 (-10%)

of measurements are written

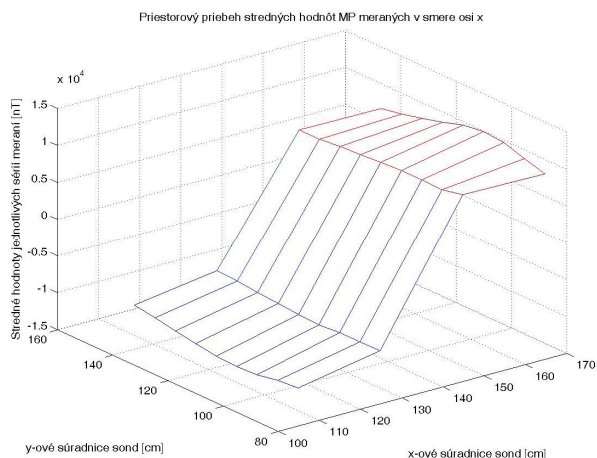


Fig. 6. Space characteristics of MPM 20 magnetic aura measured in the direction of axis x

### B. Measurement in the direction of axis y

During this measurement, probes were positioned in parallel with y axis in chosen coordinate system

In following table, meteorological conditions for this series of measurements are written

TABLE III  
METEOROLOGICAL CONDITIONS FOR 15.10.2008

Date	Dew point [°C]	Atmosph. pressure [hPa]	Medium external temp. [°C]	Medium internal temp. [°C]	Ozone [Dobs. unit]
15.10.	9	1019	14	15	270 (-12%)

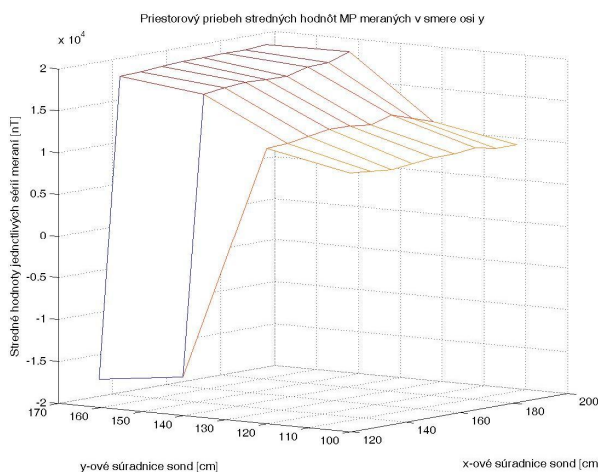


Fig. 7. Space characteristics of MPM 20 magnetic aura measured in the direction of axis y

### C. Measurement in the direction of axis z

During this measurement, probes were positioned in parallel with z axis in chosen grid system and measurement was done in two series, where in first experiment, base with probes was moved along y axis and probes were positioned in parallel with axis x. In second series, base with probes was

TABLE IV  
METEOROLOGICAL CONDITIONS FOR 24.10.2008

Date	Dew point [°C]	Atmosph. pressure [hPa]	Medium external temp. [°C]	Medium internal temp. [°C]	Ozone [Dobs. unit]
24.10.	3	1028	7	8	294 (+3%)

moved along axis x and probes were positioned parallel with axis y.

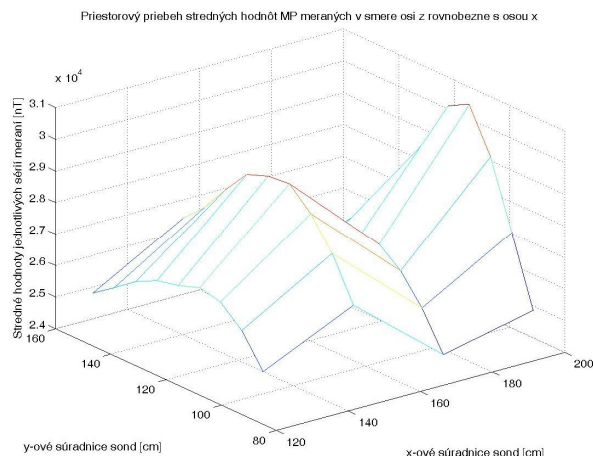


Fig. 8. Space characteristics of MPM 20 magnetic aura measured in the direction of axis z in parallel with axis x

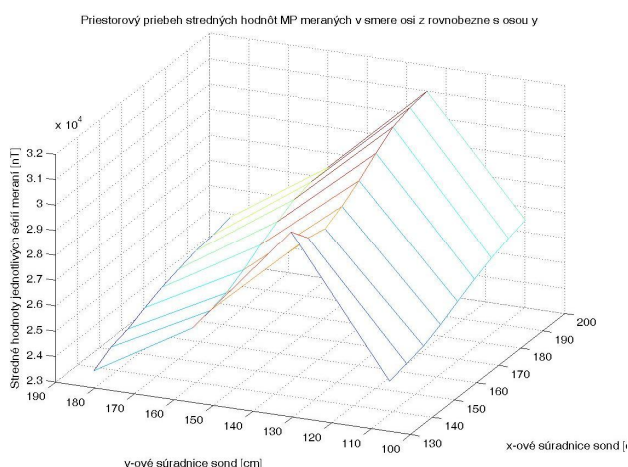


Fig. 9. Space characteristics of MPM 20 magnetic aura measured in the direction of axis z in parallel with axis y

In previous table, meteorological conditions for these series of measurements are written

From these pictures can be seen, that character of magnetic aura of MPM 20 is different in constituent components. Component of magnetic aura can be horizontal, or vertical, where horizontal components are in direction of axis x and y and vertical component is in direction of z axis.

Measurements show, that the vertical component considerably exceeds horizontal components, which is best seen in spatial graphs [8].

Described measurements were done in standstill regime of MPM 20 motor. In close future another measurements to determine the most sensitive region and component in vicinity of MPM 20 are planned. Based on the chosen region it will be possible to perform measurements in this region during working regimes of motor MPM 20 which are cold overspeed and normal working regime.

## IV. CONCLUSION

Measured results confirmed, that MPM 20 has its magnetic aura which is variable and at the same time show also the considerable difference between vertical and horizontal

components of magnetic field. One of the aims of research of magnetic aura is its future use in area of situational control of MPM 20 or supplementing the situational framework for possible improvement, or more precise control of MPM 20.

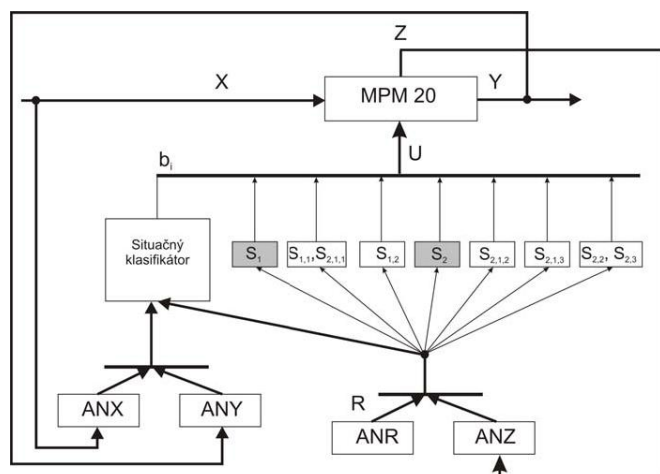


Fig. 10. Algorithm scheme for MPM 20 situational control [1]

This situational framework is more detailed described in works [1, 5] and is depicted by following structure (Fig. 10).

Based on existing experiences it can be presumed, that more precise knowledge about shape and behaviour of magnetic aura can really improve the quality of MPM 20 control. More precise description of specific design of improvement of control regimes will require other more detailed analysis of relations between certain events and processes.

#### ACKNOWLEDGMENT

The work was supported by projects: VEGA no. 1/0394/08 – Situational control algorithms and large scale systems modeling.

#### REFERENCES

- [1] Andoga R.: Hybridné metódy situačného riadenia zložitých systémov., Doktorandská dizertačná práca, FEI TUKE, 2006, pp.120.
- [2] Čopík J.: Monitorovania magnetického poľa v záujmovom priestore., Doktorandská dizertačná práca, LF TUKE, 2005, pp.135.
- [3] Čopík J.: Autonómne meranie vybraných fyzikálnych polí., Písomná práca k dizertačnej skúške, Vojenská letecká akadémia gen. M.R.Štefánika, Košice, 2002, pp.49.
- [4] Dendis T.: Letecké prístroje I., Alfa, Bratislava, 1989, pp.194.
- [5] Főző L.: Využitie matematického modelu rovnovážneho a nerovnovážneho chodu motora MPM20 pri návrhu algoritmu riadeníav každom čase, Doktorandská dizertačná práca, FEI TUKE, 2008, pp.142.
- [6] Hudák J.: Magnetometria, vybrané problémy vývoja a využitia, Habilitačná práca, Vojenská letecká akadémia gen. M.R. Štefánika, Košice, 1999, pp.107.
- [7] Kabát J., Madarász L., Modrovičová J.: Introduction to problem, basis and measurement of magnetic aura of turbojet aircraft engines, In: *Computational Intelligence and Informatics: Proceeding of the 9<sup>th</sup> International Symposium of Hungarian Researchers*, November 6-8, Budapest, ISBN 978-963-7154-82-9, 2008, pp. 379-386.
- [8] Modrovičová J.: Štúdia možnosti vplyvu magnetickej aury leteckého motora MPM 20 na jeho situačné riadenie., Písomná práca k dizertačnej skúške, FEI TUKE, Košice, 2009, pp.85.
- [9] MODROVIČOVÁ, Jana - MADARÁSZ, Ladislav - KABÁT, Ján: Magnetic aura study of small turbojet engine MPM 20. In: *SAMI 2009 : 7th International Symposium on Applied Machine Intelligence and Informatics : January 30-31, 2009, Herľany, Slovakia. [S.l.] : IEEE, 2009. s. 25-28. ISBN 978-1-4244-3802-0.*
- [10] [http://www.physics.sjsu.edu/becke/physics51/mag\\_field.htm](http://www.physics.sjsu.edu/becke/physics51/mag_field.htm)



# Basic de/composition of Time Basic Nets

<sup>1</sup>Attila N.Kovács, <sup>2</sup>Marek Výrost

<sup>1</sup>Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

<sup>2</sup>No institution, Slovak Republic

<sup>1</sup>attila.n.kovacs@tuke.sk, <sup>2</sup>marvy@reacode.com

**Abstract**—Systems, whose overall correctness depends on time, are called time-critical systems. To model these kind of systems a powerful formalism called Petri net can be used. Several extensions of Petri nets that are dealing with time have been proposed (i.e. Timed Petri Nets, Stochastic Petri Nets). This paper deals with high-level Petri nets called Environment Relationship nets and their special type called Time Basic nets. A possible time reachability problem solution is presented using a so called de/compositional approach.

**Keywords**—De/compositional approach, Environment Relationship Nets, Petri Nets, Reachability analysis, Time Basic Nets.

## I. INTRODUCTION

Petri Nets are nowadays widely used to model and analyze various types of system, such as parallel systems, database systems and also time-critical systems. Systems whose functionalities are defined with respect to time and whose correctness can only be assessed by taking time into consideration are called time-critical systems. As an example can serve a typical computer system with a mouse input device where making two clicks on the mouse (the first and the second click at the beginning and at the end of the time interval duration 2 seconds respectively), has a quite different meaning than a double click within the time interval duration a half of second.

Solving the reachability problem for a Petri net opens the door to examine other important properties like liveness or coverability. Reachability problem belongs to the class of intractable problems and it has a tremendous complexity [4].

In this paper we introduce ER nets and TB nets that can model time-critical systems. In section V. we introduce the reachability problem for Ordinary Petri Nets and especially for Time Basic Nets where a possible solution for this problem is revealed.

## II. ENVIRONMENT RELATIONSHIP NETS

Environment Relationship Nets (ER nets) represents a very strong extension of ordinary Petri Nets [1], [6]. Tokens in ER nets are called environments, i.e., functions associating values to variables (e.g., real number, string, a pair of integers). ER nets provide a possibility to incorporate the notion of the time into the concept.

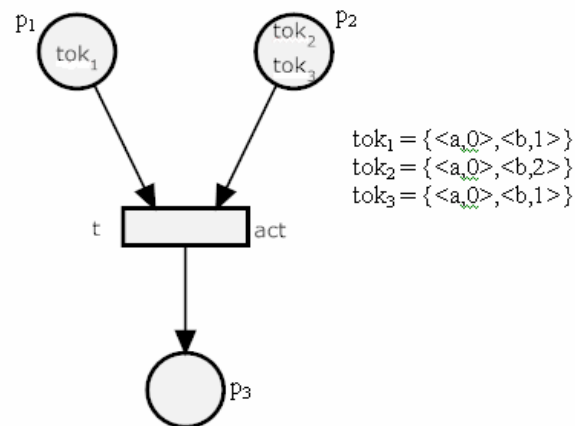
### A. Characteristics of ER Nets

An ER net is a net where:

- Tokens are environments on ID (the set of identifies) and V (the set of all values identifiers can take upon). The universal set of environments  $ENV = V^{ID}$ . Tokens and environments are used interchangeably.
- Each transition  $t$  is associated with an action. An action is a relationship

$$\alpha(t) \subseteq ENV^{k(t)} \times ENV^{h(t)} \text{ and} \\ k(t) = |\bullet t|, h(t) = |t\bullet|. \quad (1)$$

- The predicate of  $t$  is denoted by  $\pi(t)$  and it represents the projection of  $\alpha(t)$  on  $ENV^{k(t)}$ .
- A marking is an assignment of multisets of environments (*envs*) to places.
- Transition  $t$  is enabled in marking  $m$  if and only if  $\forall p_i \in \bullet t$  there is at least one token  $env_i$  and such that the tuple  $\langle env_1, \dots, env_{k(t)} \rangle \in \pi(t)$ . The tuple  $\langle env_1, \dots, env_{k(t)} \rangle$  is called the *enabling tuple* for  $t$ . Notice there can be more than one enabling tuple for  $t$  in  $m$ . The same token can belong to several enabling tuples
- A firing (of  $t$  in  $m$ ) is a triple  $x = \langle enab, t, prod \rangle$  such that  $\langle enab, prod \rangle \in \alpha(t)$ .
- The occurrence of a firing  $x = \langle enab, t, prod \rangle$  in a marking  $m$  causes a production of new marking  $m'$  for the net, obtained from the marking  $m$  by removing the enabling tuple *enab* from the preset places of  $t$ .



$$act = \{ \langle \langle p_1, p_2 \rangle, p_3 \rangle | p_1.a = p_2.a, p_1.b < p_2.b, \\ p_3.a = p_1.a, p_3.b \in \{x | p_1.b \leq x \leq p_1.b + p_2.b\} \}$$

Figure 1. ER Nets

In a natural way we define a firing sequence and the sequence of admissible markings. We may use the notation  $m \xrightarrow{x} m'$  provided that  $x = \langle enab, t, prod \rangle$  is a firing that produces  $m'$  from  $m$ . We may also define for ER nets the set of reachable markings, boundedness, liveness and other notions that can be defined for ordinary Petri Nets.

### III. TIME ENVIRONMENT RELATIONSHIP NETS

The concept of time can be incorporated into ER nets. That can be done by an assumption that each environment contains a special variable called *chronos*, whose values are of numerical type representing the *timestamp* of the environment (token). All notions introduced for ER nets remain valid, except the firing rule; the latter has to be modified. The timestamp of the token expresses the time of the environment's creation. The chronos value for the token  $env_i$  is denoted by  $env_i.chronos$ . The firing  $x$  has the structure  $x = \langle enab, t, prod \rangle$  where  $enab = \langle env_1, \dots, env_{k(t)} \rangle$ ,  $prod = \langle env'_1, \dots, env'_{h(t)} \rangle$ ,  $k(t) = |\bullet t|$ ,  $h(t) = |t \bullet|$ . In order to capture the intuitive concept of time, chronos cannot be treated as any variable.

*Axiom 1: Local monotonicity.* For any firing  $x$ ,  $env'_j.chronos \geq env_i.chronos$ , for all  $j, i$ .

*Axiom 2: Constraint on timestamps.* For any firing  $x$  there is a value denoted  $time(x)$  and such that  $env'_j.chronos = time(x)$ , for all  $0 \leq j \leq h(t)$ . The value  $time(x)$  is called the *time of the firing*.

*Axiom 3: Firing sequence monotonicity.* For any firing sequence  $s = x_1 x_2 \dots x_{|s|}$ ,

$$time(x_i) \leq time(x_j), \text{ if } i < j, 1 \leq i, j \leq |s|. \quad (2)$$

For an ER net satisfying Axiom 2 we will write  $prod.chronos$  to denote the value  $time(x)$  that all chronos have in  $prod$ . An ER net where all environments contain chronos and that satisfies Axioms 1 and 2 is called Time ER net (TER net).

For given ER net that satisfies Axiom 2, the firing sequence  $s = x_1 x_2 \dots x_{|s|}$  is time ordered, if and only if for each  $i, j$   $i < j \Rightarrow time(x_i) \leq time(x_j)$ . We will call two firing sequences  $s, s'$  equivalent if and only if  $s$  is a permutation of  $s'$ . The following result can be proven [1].

*Theorem 1:* Let E be an ER net satisfying Axioms 1 and 2. For each firing sequence  $s$  with the initial marking  $m_0$  there exists a firing sequence  $s'$  equivalent to  $s$  that is time ordered.

Given a transition  $t$  and the enabling tuple  $enab$  of the TER net, we define the *set of possible firing times*  $f$ -time:

$f$ -time( $t, enab$ ) =  $\{x \mid \langle x, prod \rangle \in \alpha(t), x = prod.chronos\}$ . Let  $s = x_1 x_2 \dots x_{|s|}$ , be a firing sequence of a TER net with the initial marking  $m_0$  and let  $m_i$  be the marking produced by the  $i$ th firing of  $s$ . The firing sequence  $s$  is *strong* if and only if it is time ordered and for each  $t' \in T$  and for each  $m_i$  ( $1 \leq i \leq |s| - 1$ ) there exists no tuple  $enab'_i$  for  $t'$  in  $m_i$  such that  $time(x_{i+1}) > \sup(f$ -time( $t, enab'_i$ )). Another axiom can

be stated to hold in TER nets.

*Axiom 3':* All firing sequences are strong.

TER net that satisfies Axioms 1, 2 and 3' is called the *strong* TER net (STER net).

The modeling power of TER nets is such that covers all known extensions of Petri Nets including those which incorporate the concept of time [1].

### IV. TIME BASIC NETS

Time Basic nets (TB nets) are a particular case of TER nets. When we assume that the only types of tokens in TER nets are chronos then we get TB nets. TB nets have been introduced in [1].

#### A. Time Basic Nets Definitios

- A TB net is a 6-tuple  $\langle P, T, \Theta, F, tf, m_0 \rangle$  where  $P, T$  and  $F$  are, respectively, the sets of places, transitions, and arcs of nets. The preset of transition  $t$ , i.e., the set of places connected with  $t$  by an arc entering  $t$ , is denoted by  $\bullet t$ .
- $\Theta$  (a numeric set) is the set of values (timestamps), associated with the tokens. A timestamp represents the time at which the token has been created. In the following, we assume  $\Theta$  to be the set of non-negative real numbers, i.e., time is assumed to be continuous.
- $tf$  is a function that associates a function  $tf_t$  (called time-function) with each transition  $t$ . Let  $enab$  be a tuple of tokens, one for each place in  $\bullet t$ . Function  $tf_t$  associates with each tuple  $enab$  a set of value  $\theta$  ( $\theta \subseteq \Theta$ ), such that each value in  $\theta$  is not less than the maximum of the timestamps associated with the tokens belonging to  $enab$ .

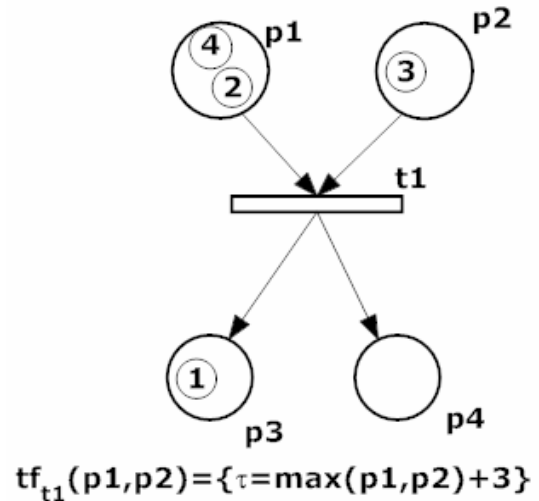


Figure 2. Time Basic Net

#### B. Enabling Tuple, Enabling, Firing Time, Enabling Time

Given a transition  $t$  and a marking  $m$ , let  $enab$  be a tuple of tokens, one for each input place of transition  $t$ . If  $tf_t(enab)$  is not empty,  $enab$  is said to be an enabling tuple for transition  $t$  and the pair  $x = \langle enab, t \rangle$  is said to be an *enabling*. The triple  $y = \langle enab, t, \tau \rangle$  where  $\langle enab, t \rangle$  is an enabling and  $\tau \in tf_t(enab)$ , is said to be a firing.  $\tau$  is said to be the firing time. The maximum among the timestamps associated with

tuple  $enab$  is the enabling time of the *enabling*  $\langle enab, t \rangle$ .

The dynamic evolution of the net (its semantics) is defined by means of firing occurrences, which ultimately produce firing sequences. Markings represent the states of the modeled system; transitions represent events; and firing sequences represent evolution of the modeled system.

The following axioms must hold in TB nets: time never decreases; if the system does not stop, time eventually progresses. More axioms for TB nets can be found in [2].

### C. Time Interval Semantics of Time Basic Nets

Interval semantics of TB nets give us the opportunity to assign to each token a time interval (TI) in which the token can be created [3]. Using TI instead of timestamps gives us a bigger modeling power. Any token (chronos)  $\tau$  in TI  $\tau = [\tau_i, \tau_a] \subseteq \mathbb{R}^+$ . In TI semantics we replace any enabling tuple  $enab = (m(p_1) \dots m(p_{\bullet_i}))$  with a corresponding collection of TIs that is called *Time Interval Profile* (TIP). Besides the set operation  $\cap, \cup, ()^c$  a new operation “+” is defined. Given a constant  $c \in \mathbb{R}^+$  and TIs  $\tau'$  and  $\tau''$  we have:  $c + \tau = [\tau_i + c, \tau_a + c]$ ,  $c \cdot \tau = [\tau_i \cdot c, \tau_a \cdot c]$ ;  $\tau' + \tau'' = \tau \Leftrightarrow \tau = [\tau_i, \tau_a]$ ,  $\tau' = [\tau'_i, \tau'_a]$ ,  $\tau'' = [\tau''_i, \tau''_a]$ ,  $\tau_i = \tau'_i + \tau''_i$ ,  $\tau_a = \tau'_a + \tau''_a$ .

Given TB net  $N_0 = (P, T, \Theta, pre, post, tf, q_0)$ , then  $tf_i(enab)$  has for given  $enab$  the unique representation

$$tf_i(enab) = \tau en + tf_i(0) \quad (3)$$

where  $\tau en$  is a TI that depends on  $enab$ ,  $tf_i(0)$  is a TI that does not depend on  $enab$ . To put it another way, any  $t$ -generated TI  $\tau_i$  can be represented as a sum of two TIs:  $\tau en$  - the determinate TI that depends on TIP  $enab$  in question and on  $t$  (or  $tf_i$ ) and a constant TI  $tf_i(0)$ , which depends only on the structure of the TB nets in question.

## V. TIME REACHABILITY ANALYSIS OF TIME BASIC NETS

Given Petri net  $N_0 = (P, T, pre, post, m_0)$  and a  $k$ -dimensional nonnegative integer vector  $q$ ; the problem whether  $q \in \mathfrak{R}(N_0)$  is called (the instance of) the reachability problem (RP) of PN  $N_0$  for state  $q$ , where  $\mathfrak{R}(N_0)$  is the set of all reachable markings in  $N_0$ .

In Time Basic Nets we not only ask, whether a specified marking is reachable, but we also want to if it is reachable in specified time (time interval). The complexity of the solution for the reachability problem in the worst case for Ordinary

Petri nets is tremendous:  $O(2^{b^{k+2} \cdot k^2})$ , provided that  $k = \text{card } P$ ,  $P$  is the set of places of PN  $N_0 = (P, T, pre, post, m_0)$  and  $b \geq 0$  is some constant. It turns out that only possible way to solve the RP is to use a so called de/composition of the ordinary PN. Given a PN  $N = (P, T, pre, post, m_0)$  we are looking for some method how to split  $N$  into two subnets  $N_i = (P_i, T_i, pre_i, post_i, m_{0i})$   $i = 1, 2$  such that there is clear relation between languages of Petri nets  $L(N), L(N_i)$ , reachable states  $\mathfrak{R}(N), \mathfrak{R}(N_i)$ , fsa of type  $M_w M, M_i$  respectively and also there would be splendid if

$$q \in \mathfrak{R}(N) \Leftrightarrow q'_1 \in \mathfrak{R}(N_1) \wedge q'_2 \in \mathfrak{R}(N_2) \quad (4)$$

provided that  $q'_1, q'_2$  are corresponding images of  $q$  under the operation that we are looking for. The operations that reliably do this are called T-JUNCTION, P-JUNCTION and PT-JUNCTION [4], [5].

In the Fig 3. the T-JUNCTION method was applied for the net  $N_0$ . The reachability problem is separately computed for the nets  $N_1, N_2$  and the results are combined to get the solution for the whole net.

Using a T-JUNCTION method (cutting the net through transitions) in the case of Time Basic Nets also the transition

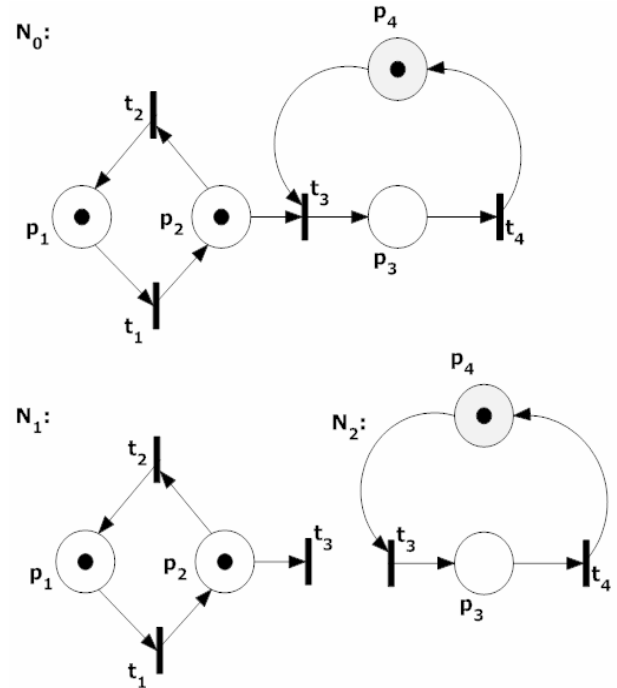


Figure 3. Ordinary Petri Net T-de/composition

predicate de/composition is needed [6]. It can be proved that each basic predicate can be de/composed to predicates

$$tf_i'(enab_1) = A', \quad tf_i''(enab_1) = A'' \quad \text{for the net } N_1 \quad (5)$$

$$tf_i'(enab_2) = B', \quad tf_i''(enab_2) = B'' \quad \text{for the net } N_2, \quad (6)$$

$$enab_1 \cup enab_2 = enab.$$

A transition predicate  $tf_i(p_i, p_j) = \tau = (\max(t_i, t_j) + a)$ , where  $a$  is some constant, can be de/composed to predicates  $A' = \tau_i + a$ ,  $A'' = \tau_i + a + R^+$  and  $B' = \tau_j + a$ ,  $B'' = \tau_j + a + R^+$ , where  $R^+ = [0, +\infty]$ .

## VI. CONCLUSION

Environment Relationship Nets and Time Basic Nets are a powerful Petri Nets for modeling and analyzing time-critical systems. Analysis can reveal errors or faulty states in the modeled system. To reveal these errors the solution for the reachability problem plays a profound role in this case. It was shown, that the only possible way how to practically deal with this problem is to use the de/compositional approach. This approach can be nowadays used just partly in the case of Time Basic Nets. Future work will be oriented to the de/composition of all kinds of transition predicates for formerly mentioned nets.

#### ACKNOWLEDGMENT

The work was supported by project: VEGA no. 1/3140/06 “De/compositional Design and Analysis of Discrete Systems Based on Petri Nets” and NATO Linkage Grant ES CLG 982698 “Verification of Complex Networking Protocols”.

#### REFERENCES

- [1] C. Ghezzi, D. Mandrioli, S. Morasca, M. Pezzé, “A unified high-level Petri net formalism for time critical systems (Book style with paper title and editor),” IEEE Transition on Software Engineering, 2nd ed. vol. SE-17, 1991, pp. 10-150.
- [2] M. Felder, C. Ghezzi, M. Pezzé, *Analyzing refinements of state based specification: the case of TB nets*. New York: Springer-Verlag, 1994, ch. 7.
- [3] Š. Hudák, Time Interval Semantics of TB Nets, Proceeding of international conference RSEE'96, Romania, 1996, pp. 1-12.
- [4] Š. Hudák, “Reachability Analysis of Systems Based on Petri Nets (Book style),” Košice, elfa s.r.o., 1999, pp. 47-180.
- [5] Š. Hudák, S. Šimoňák, Š. Korečko, A. N.Kovács, Formal Specification and de/composition approach to design and analysis of discrete systems, Košice, elfa, 2007. pp. 8-46.
- [6] J. Bača, Š. Hudák, De/compositional Time Reachability Analysis, Proceedings of the 6th International Conference, Oradea - Felix Spa, Romania, 24.-26.5.2001, Oradea - Felix Spa, 2001, pp. 60-65.

# Modification of DSR to implement SSV to the Mobile Ad-hoc Network

Ján PAPAĽ

Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

jan.papaj@tuke.sk

**Abstract**— A mobile ad-hoc network (MANET) is infrastructure-less, self-organized networks that either operate autonomously or as an extension to the wired networking infrastructure. MANET is expected to support new QoS and security-based applications. Nowadays, the QoS and security are very important areas of research in MANET, but are considered separately. No protocol or system is designed to integrate QoS and Security to the integrated system. This paper examines the integration of Security Service Vector to MANET. We design a new user-based model to interaction between QoS, Security and routing protocol. The integration provides the user with new abilities to configure QoS and security requirements.

**Keywords**— DSR, MANET, QoS, Security, Security Service Vector.

## I. INTRODUCTION

A mobile ad-hoc network (MANET) is a collection of mobile nodes that are self-configuring and capable of communicating with each other, establishing and maintaining connections as needed. Nodes in MANET are both routers and terminals. These networks are dynamic in the sense that each node is free to join and leave the network in a nondeterministic way.

The notion of QoS is a guarantee provided by the network to satisfy a set of predetermined service performance constraints for the user in terms of the end-to-end delay statistics, available bandwidth, probability of packet loss, and so on[1]. In literature, the researches on QoS support in MANETs include *QoS models*, *QoS resource reservation signaling*, *QoS routing* and *QoS Medium Access Control(MAC)*[1].

Security is an area that has been studied since the beginning of computing, and some aspects, such as cryptography, were studied even earlier than that. The main goals of security requirements are following: *Confidentiality*, *Authentication*, *Availability*, *Integrity* and *Non-repudiation*. In literature, the researches on security support in MANETs include *Secure routing*, *Key management* and *Intrusion Detection System*[2].

This article deals with basic modification of standard routing process in MANET environment. The Dynamic Source Routing protocol (DSR) is used to integrate Security Service Vector to the new designed model. The integration provides the user with new abilities to configure QoS and security requirements via routing protocol. This implies that users have

opportunity to configure their own level of security and QoS parameters for services.

## II. SECURITY SERVICE VECTOR (SSV)

Concept of Security service vector(SSV) is introduced in [1]. Security vector (SV) was proposed to determine a number of customizable Security Services (SSs) with choices of customizable Service Degrees (SDs). There are two communication phases taken place for data transmission: **probing phase** and **data transmission**.

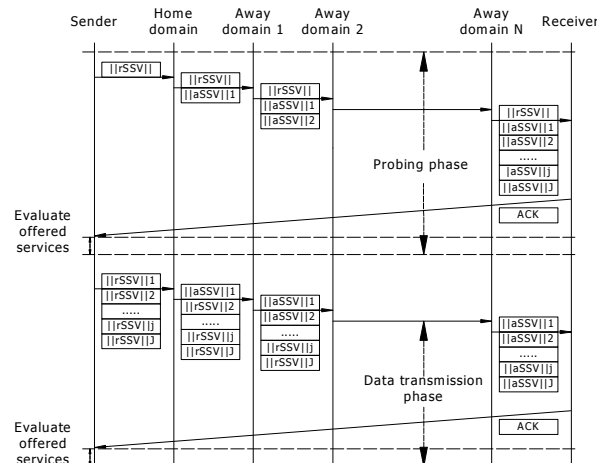


Fig. 1. The transmission diagram of SSV

**The probing phase** happens during a connection establishment and the data phase starts after the connection has been set up. During the probing phase, the sender who wants to exercise security service options sends the probing packet through all domains along the path.

The probing packet verifies whether satisfied security services can be offered along with sufficient resources for the following data packets. This probing packet contains the same requested SSV ( $rSSV$ ) for every domain. The  $rSSV$  portion in the probing phase is denoted as

$$\|rSSV\| = \{(SS^1, SD_m), \dots, (SS^N, SD_m), [data\_length]\} \quad (1)$$

Any edge router performs the following basic tasks: verifying the sender's identity; examining the  $rSSV$ ; verifying whether the sender has a privilege to request the services; checking available resources; writing down its  $aSSV$  portion; and forwarding the probing packet to the next hop. The receiver replies with an ACK packet containing all available SSV ( $aSSV$ ) portions to the sender.

The sender then evaluates all services offered and concludes whether to proceed to the data phase or to drop this connection and try again later. At the end of the probing phase, the querying user retrieves information from all *aSSV* portions carried in the *ACK* packet is denoted as

$$\|aSSV\| = \bigcup_{j=1}^J \|aSSV\|_j \equiv$$

$$(SS^1, SD_m), \dots, (SS^N, SD_m) \{ (SS^1, SD_m), \dots, (SS^N, SD_m),$$

$$[delay, time\_process, cost]estimated \}_1, \dots,$$

$$(SS^1, SD_m), \dots, (SS^N, SD_m) \{ (SS^1, SD_m), \dots, (SS^N, SD_m),$$

$$[delay, time\_process, cost]estimated \}_J$$
(2)

Satisfied with the evaluation result, the user starts the **data transmission phase** during which the data flow is attached with security-related information and sent through the network. In other words, the *rSSV* portions, one for each intermediate router, are attached into each data flow [4]. The *rSSV* portions in the data transmission phase are denoted as

$$\|rSSV\| = \bigcup_{j=1}^J \|rSSV\|_j \equiv \{ (SS^1, SD_m), \dots, (SS^N, SD_m) \}_1, \dots,$$

$$\{ (SS^1, DS_m), \dots, (SS^N, SD_m) \}_J$$
(3)

Upon an arrival at each router, a router picks up its associated *rSSV* portion and executes the security services requested individually. After the security services were served, each router records the results by replacing the corresponding *rSSV* portion with the *aSSV* portion to report the querying user.

### III. THE SSV ARCHITECTURE IN WIRED NETWORK

The SSV architecture proposed to cooperation between QoS and security mechanisms in wired network [4].

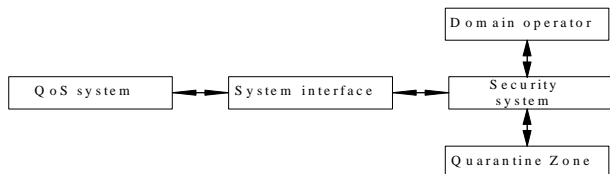


Fig. 2. The SSV architecture for wired network

QoS system controls numbers of requirements and incoming traffic. When it receives very large amount of incoming traffic from the same source or same destination, system alerts the security system to verify whether the network is being under attack. Report log is stored for subsequent analysis. Quantitative zone is deployed to keep some packets as suspicious attacking packets while waiting for the user's confirmation.

### IV. PROCESS OF INTEGRATION TO THE MANET

We design new architecture to cooperation between users, services, QoS, security and routing mechanisms via SSV (see Fig.3).

In this model, the user will specify requirements on QoS and security via SSV and then the routing protocol selects an appropriate route to the destination, and consequently the network can provide the user-requested services. The 'appropriate' means that all nodes around the selected nodes can fulfill selected requirements to QoS and security

requirements.

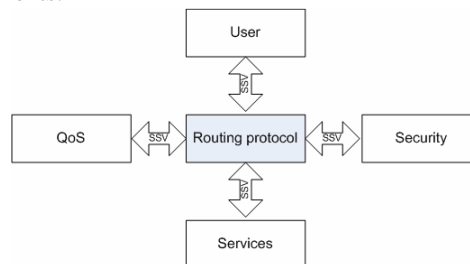


Fig. 3. The SSV architecture for MANET

The SSV will be implemented into MANET network through Dynamic Source Routing (DSR). DSR protocol is a simple and robust routing protocol designed for use in multi-hop wireless ad-hoc networks of mobile nodes [6]. Process of integration SSV to MANET is divides to following phases:

- **Probing phase (route discovery RREQ)**
- **Data transmission phase (route replay RREP)**

In first phase, the user (source node) defines its own requirements for QoS and security. The information is collected to the DSR header [6]. In second phase, the user (source node) who wants to make security-related service with another user (destination node) or server sends the modified route request packet to the all-intermediate nodes along the source node (Fig. 4).

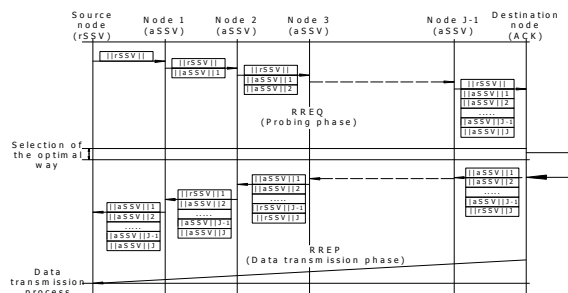


Fig. 4. Implementation of SSV to the routing process

Request packet contains requirements specified for QoS and security. If intermediate node is not the destination node, added is information about possibilities of provisioning requested services. This information is added to the relevant aSSV and then is sent to another intermediate node. This process is called route discovery and corresponds with Probing phase. This stage is graphically illustrated on Fig.5. Process of collecting aSSV information is executed till the destination node is found. When destination node is found, route is written to the source node route cache memory and destination node rewrites all aSSV information into the rSSV information. This packet is then called ACK packet. The ACK packet is sent back to the source node.

### V. SIMULATION AND RESULTS

OPNET simulator is used to simulation of integration process SSV to the MANET. We design simple model to implementation of SSV via DSR. For simplify we simulate only probing phase (RREQ). This process is described in Fig.5. We create project with two scenarios. All scenarios contain 10 nodes and simulate simple voice services. First scenario is standard MANET network with DSR and second is MANET with modified DSR.

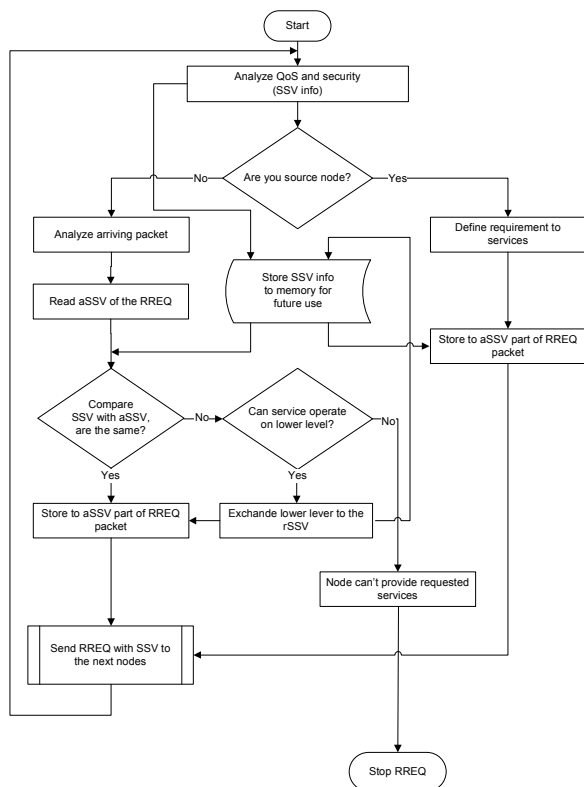


Fig. 5. Modified routing process for MANET

Delays caused by adding process of aSSV or rSSV was an observed parameter of simulation. Total delay of MANET network illustrates Fig.6. As we can see integration of SSV to DSR do not influence the total delay of the tested MANET network.

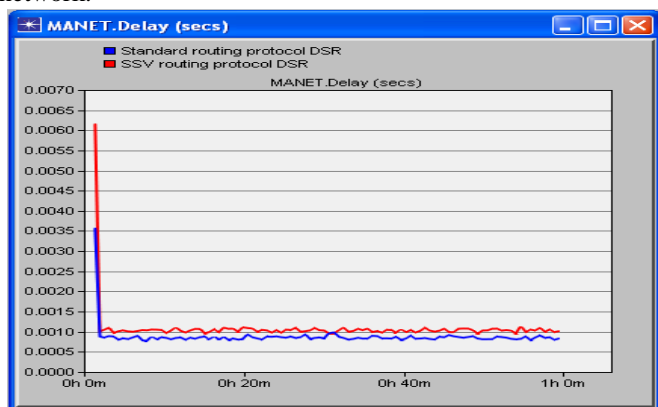


Fig. 6. Total delay of MANET

On Fig.7 we can see how the SSV affect process of routing in source node and other nodes. Processing time is time necessary to execute all SSV processes to integrate to the DSR on the nodes.

### VI. CONCLUSION

Today, QoS and security related mechanism don't provide user's ability to change level of requested parameters or services. In this article we introduce new user, QoS and security-related model (Fig.3). Model provides the user ability to specify its own requirements on QoS and security. All components are interactive cooperated via SSV.

We show the integration do not extremely increase total delay of sending probing packet between source and destination node as we can see on the fig.6-7. Next part of our work will be dedicated to implement SSV as a part of new model.

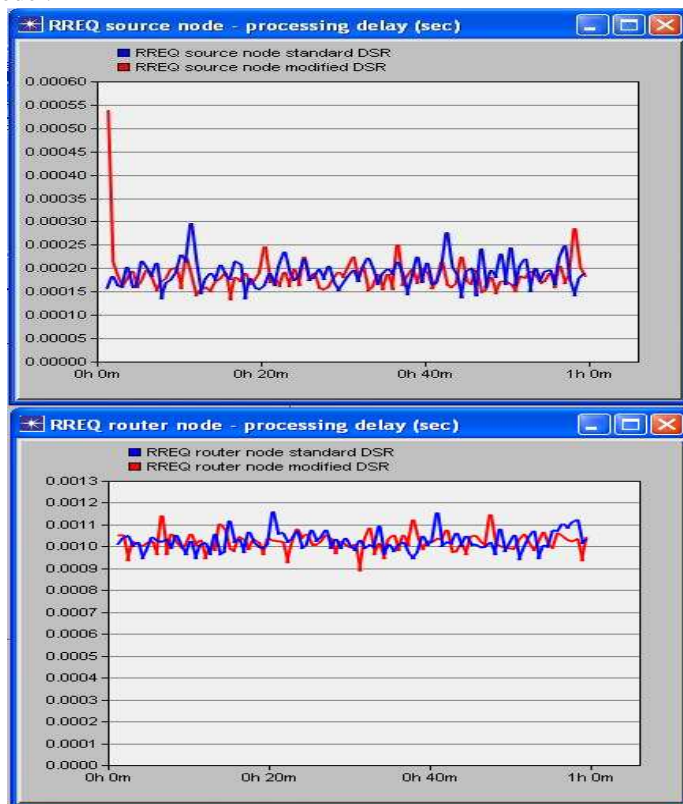


Fig. 7. Processing time in MANET nodes

### ACKNOWLEDGMENT

Research described in the paper was financially supported by VEGA No.1/4054/07, also by COST 2100 - Pervasive Mobile & Ambient Wireless Communications and INDECT.

### REFERENCES

- [1] J. Yang, J.Ye, S. Papavassiliou, "A new differentiated service model paradigm via explicit endpoint admission control", in Proc. SCC2003, Jun. 2003, pp. 299–304.
- [2] H. Yang, Y. Luo, F Ye, S W. Lu, and L Zhang: Security in mobile ad hoc networks: Challenges and solutions (2004). IEEE Wireless Communications. 11 (1), pp. 38-47. Postprint available free at: <http://repositories.cdlib.org/postprints/618>
- [3] P. Sakarindr, N. Ansari, R. Rojas-Cessa, S. Papavassiliou, "Security-enhanced quality of service (SQoS) networks", IEEE Sarnoff Symposium on Advanced in Wired and Wireless Communications, April 2005, pp. 129-132.
- [4] P. Sakarindr, N. Ansari, R. Rojas-Cessa, S. Papavassiliou, "Security-enhanced quality of service (SQoS) networks: a network analysis", IEEE Military Communications Conference, October 2005.
- [5] D. Johnson, D. Maltz, and J. Broch, "DSR: The Dynamic Source Routing Protocol for Multihop Wireless Ad Hoc Networks in Ad Hoc Networking", Addison-Wesley, 2001, pp. 139–172.
- [6] J.Papaj, L.Doboš, A.Čizmar, "Security service vector in MANET", In: AEI '2008: International Conference on Applied Electrical Engineering and Informatics: September 8-11, Greece, Athens 2008. Košice, FEI TU, 2008. p. 130-138. ISBN 978-80-553-0066-5.

# Distributed GPGPU

Ján PERHÁČ

Dept. of Computer and Informatics, FEI TU of Košice, Slovak Republic

jan.perhac@tuke.sk

**Abstract**— More and more effort has been made to increase the computational power of latest generations of graphical hardware. Nowadays, this hardware represents a massive parallel computational architecture and with its latest improvements in graphic processor unit programmability the graphical hardware can be utilized not only in classic image or scene manipulation or rendering but also in advanced physics simulation and in other computational or scientific areas as an high-performance mathematical coprocessor. This fact represents one of latest possibility how to increase the overall power of computer systems not only in present times and also in the future. This paper focuses on graphics hardware utilization possibilities especially in distributed computer systems.

**Keywords**—CPU, GPU, GPGPU, distributed computer systems, CUDA

## I. INTRODUCTION

Desktop computers, notebooks, netbooks, server workstations, game consoles, mobile devices. They all are equipped not only with central processor unit (CPU), but also with a high performance graphics processor unit (GPU) chip, which comes mostly with large dedicated memory. The CPU is designed to run the operating system of any device and applications programs of any kind, written in many computer (high level) languages like C, C++, C#, JAVA, FORTRAN, PYTHON and so on because of its multipurpose architecture design. Compared to CPU the GPU has more specialized design, which better provides realization of graphical tasks like rendering of 3D scenes, image rasterization and other transformations. For many GPU generations the functionality or programmability of GPU processors were very limited. In nowadays, the GPU processor are able to compute ten millions of vertices and to rasterize hundreds of millions or more fragments per second. The computational time for GPU is significantly shorter the computational time of the same problem on a CPU processor. But GPU processors are not able to perform any kind of program for general purpose tasks, like the CPU processor can. For many years the GPUs were used only for acceleration of processing of some parts of graphical calculation.

## II. GPU PROGRAMMING OVERVIEW

The classical architecture of GPUs was changed dramatically towards vertex nad fragment shader architecture (SA) which was enhanced into unified shader architecture (USA). SA and USA allows now to perform also non-

graphical operations on graphical hardware with an ease. The emerge of SA and USA also led to a new computer branch called General-Purpose computing on Graphics Processing Units (GPGPU), which utilize the computational power of specialized high-performance computer hardware based on SA or USA GPU processors. This new branch, about three years of age, gave impulse for creation of new programming languages, suitable especially for GPU processors. Some of these GPU based languages are: High Level Shader Language (HLSL), OpenGL Shading Language (GLSL), C for graphics (Cg), Compute Unified Device Architecture (CUDA), Close to Metal (CTM) and many others. Graphical processor is designed especially for processing of graphical operations and so its functionality is limited for operations execution and programmability. Considering the construction character of graphical hardware, the GPU is effective only in solving of problems, which can be executed in parallel. This means, that the GPU processor is able to concurrently process large amount of independent vertices and fragments. Basically, the GPU processor is a stream processor, which is able to execute one kernel in parallel on top of many data concurrently. The stream represents a dataset which requires identical computations. Streams are creating data parallelism. A kernel is a function, which is applied on every single element in the stream. Each vertices and fragments represents elements of stream, on which vertex and fragment of unifies shaders are executing given kernel.

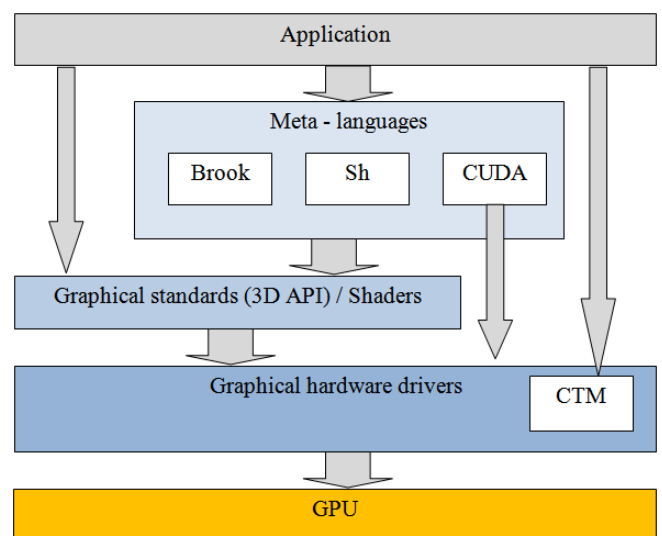


Fig. 1 Existing possibilities of programming GPGPU applications



### III. DISTRIBUTED GPGPU, D-GTS ARCHITECTURE

In present, the GPGPU research is focused primary on its utilization in classic desktop PC systems. The numerical power increase in such systems is possible right now only by using technologies like SLi and Crossfire, which increases the power of graphical adapters only on the level of graphical operations and not on the level of general purpose operations by using GPU processors. So to increase the overall numerical power in GPGPU operations it is needed to think about interconnection of desktop PC systems into distributed computer systems with taking the focus of GPGPU related problems. It is needed because similar to CPU processors, the increase of numerical power, programming possibilities and raising of specific development and administrative tools will result in near future in usage of GPU processors as mathematical coprocessors within large distributed computer systems. The research on Department of Computers and Informatics on Technical University of Košice is focusing on the design and implementation of complex solution of graphical distributed computer system with the aim on GPGPU. The proposed architecture is called D-GTS and is depicted on figure 2. This research is part of grant projects VEGA 1|4071|07 and APVV 0073-07.

The motivation / reason for the usage of latest graphical hardware and proprietary language in the design of distributed D-GTS system is the fact, that the next specification of this system will be dramatically simplified. Also the abandon of CUDA to graphical standards allows simplified administration and also the possibility to be operational under different operating systems (Windows, Linux, Mac OSX) and also simplified mechanisms implementation of operations execution using the unified computational units.

The proposed D-GTS architecture consists of next components:

- **Nodes** – there are three types of nodes. Nodes waiting for computations, requests processing node and computational nodes.
- **Entities** – there are three types of entities. Entities for describing of node character, communication messages and execution commands
- **Fronts** – there are two types of fronts. Nodes waiting for computations front and tasks front.
- **Lists** – there are two types of lists. List of all system elements and list of all system tasks.

The overall architecture is depicted on fig.2.

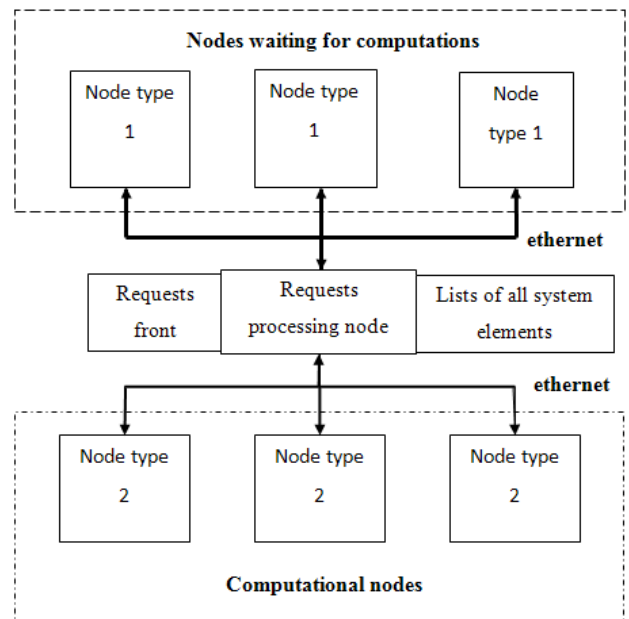


Fig. 2 Architecture D-GTS

The common element of all nodes in D-GTS distributed system is the middleware. This program framework is used in this design for basic connectivity of every single node between each other in the meaning of communication, message passing and commands passing.

**Realization of node waiting for computation** – this node represents node in term of GPGPU application, so its highest layer of realization represents the GPGPU application. Under this layer is the CUDA metaprogram. This block represents encapsulation of basic CUDA operations for execution on the graphical hardware in the form of script, which will be compiled on the computational node. This node is shown on the fig. 3.

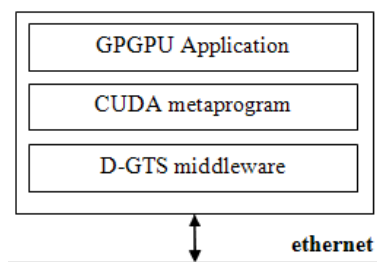


Fig. 3 Realization of node waiting for computation

**Realization of computational node** – this node represents node in the term of data – parallel computations, so its lowest layer is represented by graphical hardware. Next there is the D-GTS middleware to ensure the connectivity between node and CUDA compiler, which main task is to compile CUDA metaprogram into binary form which is able to be executed on the graphical hardware. The highest layer is represented by the D-GTS node control application, which represents main coordination with requests processing node. This node is shown on fig. 4.

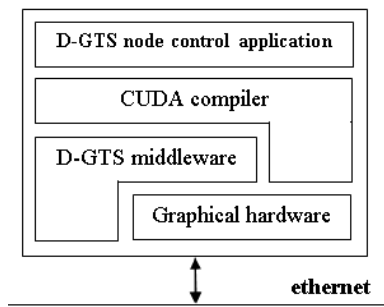


Fig. 4 Realization of computational node

**Requests processing node** – this node represents main system node and contains mechanisms for administrations of nodes which are waiting for computations and mechanisms for administration of requests front and list of all system elements. Next there is the D-GTS middleware for possibility to communication with each system element. This node receives and registers requests for graphical hardware from nodes waiting for computations. If one of the computational node is freed than the CUDA script sent from by D-GTS node control application. This node compiles given script and sends it to computational node. The requests processing node is depicted on fig. 5.

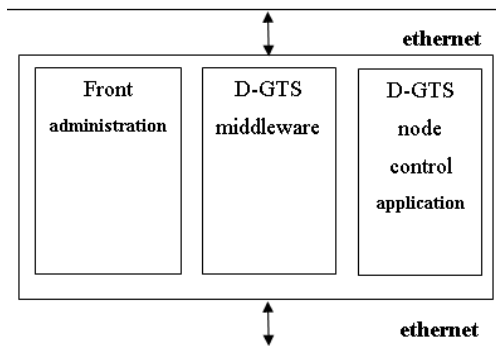


Fig. 5 Requests processing node

#### 4. CONCLUSION

The GPGPU problematic has a promising outlook. GPU processors are even now powerful enough coprocessors to main CPU in areas like physics, artificial intelligence. The need of effective utilization of GPU processors potential is known also by GPU manufacturers / OS manufacturers, which are implementing some form of acceleration by graphical adapter into their products. The computational growth of these adapters will continue to rise, because the computational growth is not the main question of GPU architecture, but the question of number of computational shader units. It is assumed that all future graphical adapters will be fully programmable. That is why the thinking about utilization of graphical adapters within one distributed network is important. Distributed GPGPU will be in near future the same obviousity as distributed CPU networks are. So it is only the matter of time, who will design the most effective way for distributed solving of GPGPU problems. The research on Department of

Computers and Informatics on Technical University of Košice tries to contribute to this process with own ideas.

#### REFERENCES

- [1] Pharr M., Fernando R.: GPU Gems 2: Programming Techniques for High-Performance Graphics and General-Purpose Computation, Addison-Wesley Professional, 2005, ISBN 0321335597
- [2] Heinecke A.: A Short Introduction to nVidia's CUDA, Ferienakademie, 2007
- [3] Shirley P.: Fundamentals of Computer Graphics, AK Peters, Ltd., England, 2002, ISBN 1568811241
- [4] Vokorokos L.: Digital Computers Principles, Typotex 2004, Budapest, ISBN 9639548 09
- [5] Sobota Branislav, Perhác Ján, Szabó Csaba, Schrötter Štefan: High-resolution visualisation in cluster environment, Grid Computing for Complex Problem 2008, Bratislava, 27.10.-29.10.2008, Bratislava, Ústav Informatiky, SAV, 2008, 4, pp. 62-69, ISBN 978-80-969202-9-7

SUPPORTED BY VEGA PROJECT No. 1|4071|07

AND APVV 0073-07

# De/compositional analysis of Petri nets – a survey

Ivan Peřko

Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

ivan.petko@toryconsulting.sk

**Abstract**—The aim of this paper is to briefly survey some of the de/compositional approaches in the analysis of Petri nets. Basic ideas are presented in order to illustrate a wide variety of approaches. Instead of giving a comprehensive survey, some of the results in the de/compositional analysis are described. The focus is on analytical problems and techniques, principles of de/composition and special classes of Petri nets developed to introduce de/compositional semantics.

**Keywords**—analytical problems, de/compositional analysis, Petri nets.

## I. INTRODUCTION

Petri nets [1] (PN) are a well known formal method that allows the formal description and analysis of discrete systems. First introduced by C.A. Petri in his dissertation “Kommunikation mit automaten” in the early 1960s, the main advantages they offer are simplicity and intuitive semantics. A lot of research was devoted to their investigation and this led to a wide family of Petri nets covering many aspects of real systems. Petri nets are especially suitable for modeling systems with characteristics such as parallelism, synchronization, distributed computing, and resource sharing. Their analytical properties and the possibility of automatic verification make them very useful and interesting in the academic field and helpful in practice.

Although there are many advantages, there are some drawbacks as well. One of the main disadvantages has been their capability for de/composition. Early on, researchers noticed that a larger system involves a larger net and this is not easy to analyse. Research efforts were undertaken in order to overcome this problem and techniques were developed for de/composition and special classes of compositional Petri nets. This paper presents a brief survey of research in the field of de/composition and de/compositional analysis of Petri nets.

## II. PETRI NET ANALYSIS

### A. Analytical problems

Analytical problems are closely related to structural or behavioral properties of Petri nets. The most crucial ones [2] are *Reachability Problem (RP)*, *Liveness Problem (LP)*, *Boundedness Problem (BP)*, *Deadlock Problem (DP)*, *Model Checking Problem (MCP)*, and *Equivalence and Containment Problem (EP, CP)*.

RP is the central issue of Petri net theory. Since many problems are recursively equivalent to the RP, it is sufficient to solve the RP to answer others. This, seemingly simple task,

can be quite difficult to solve when considering models of real systems, which are often huge and/or suffer from state explosion problem (SEP). Therefore, the de/compositional approach is often the only way to solve analytical problems.

### B. Analytical techniques

Analytical techniques [3] can be basically divided into four groups:

- *Algebraic techniques* – based on the ILP (integer linear programming) concept, these provide a simple method for analysing net properties (e.g. invariant) by creating and solving equations describing net characteristics (using an incidence matrix and principles of linear algebra)
- *Structural analysis* – answers structural questions (e.g. BP)
- *State-space analysis* – the simplest and most straightforward technique describing net behaviour by the reachability and coverability graphs oriented to the RP
- *Techniques specific for the given class of Petri nets* (e.g. Timed PN)

Answering analytical problems for “sufficiently” big PN involves the de/compositional approach, even when analysing with the aid of computers. Thus the de/compositional analysis could be the fifth analytical technique that covers the four mentioned above.

## III. DE/COMPOSITIONAL ANALYSIS

In order to analyse real systems, many researchers paid attention to the de/composition of PN. The following briefly presents some of the achieved results.

### A. De/compositional approaches

#### Functional subnets

In the series of papers, Zaitsev introduced the concept of functional subnets [4], i.e. nets with input and output places with no input arcs for input places and output arcs for output places. By extending the concept to minimal functional subnets, Zaitsev proposed the algorithm of net decomposition [5] with polynomial time complexity (depending on the size of the net).

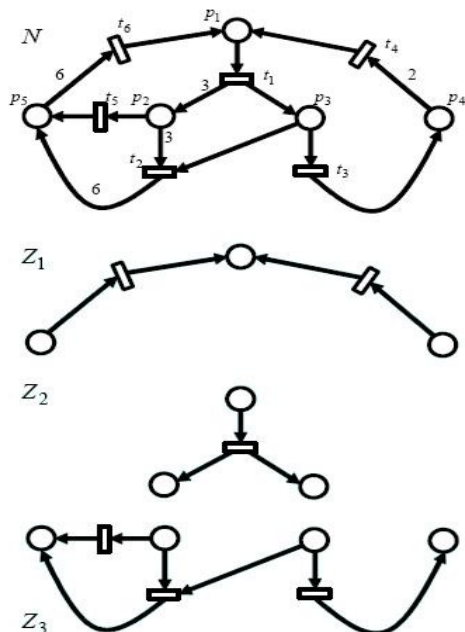


Fig. 1. Net  $N$  and its corresponding minimal functional subnets  $Z_1, Z_2, Z_3$

De/compositional analysis is oriented towards the invariant calculation [6] and structural properties of the whole net from its subnets (using composition and decomposition of functional subnets). The obtained acceleration is exponential. This approach is subsequently used for protocol verification (ECMA, TCP, Ethernet...).

*De/composition based on indexes of places*

Another approach was chosen in Zeng’s works. Zeng and his colleagues developed a polynomial time algorithm [7] based on index function ( $f: P \rightarrow \mathcal{N}$  such that if  $\forall p_1, p_2 \in P: (\bullet p_1 \cap \bullet p_2 \neq \emptyset) \vee (p_1 \bullet \cap p_2 \bullet \neq \emptyset) \Rightarrow f(p_1) \neq f(p_2)$ ) indexing the places of nets such that the resulting subnets are all always simple nets (S-nets:  $\forall t \in T_{S-net}: |t| \leq 1$ ). This decomposition technique is not unique and depends on the index function.

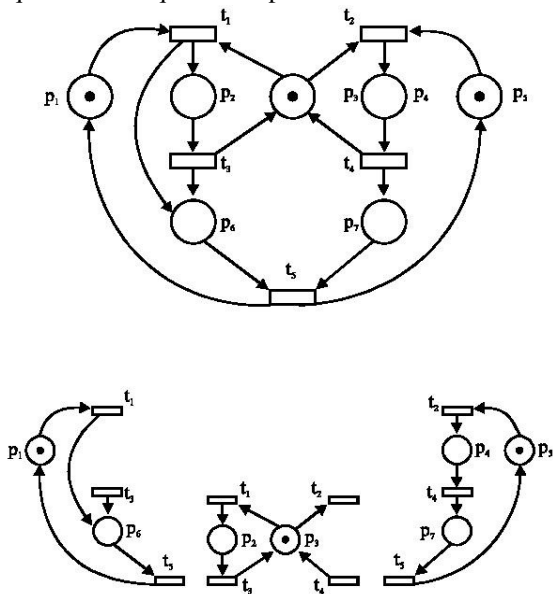


Fig. 2. Petri net and its S-nets obtained by de/composition based on index places

Zeng subsequently analyses in detail the relationships between the reachable states, languages, liveness and invariant fairness [8] in the original Petri net and its subnets and provides the sufficient and necessary conditions to keep the reachable states and languages invariant during the decomposition.

Zeng introduces the synchronous composition based on the same concept (index function) [9] in order to get structure-complex nets and obtain its language expressions from its subnets (S-nets), investigating the language relationships between net and its subnets.

*De/compositional analysis with respect to solving the reachability problem*

In [2] Hudák considers the problem of splitting a net  $N$  into two subnets such that the RP can be solved separately for the resulting subnets. The author introduces three types of decomposition: T-junction (splitting with respect to (synchronic, common) transitions), P-junction (splitting with respect to (common) places) and PT-junction (a combination of P-junction and T-junction) and decomposition algorithms. It is shown how the decomposition affects the  $M_W$  automaton (the finite state automaton constructed from PN in order to solve the RP) and thus solves the RP. The author also deals with the construction of the corresponding  $M_W$  automaton for subnets obtained by the decomposition from that for the original PN, and shows how the RP can be solved in a de/compositional manner.

*Composition operations on High-Level PN (HLPN)*

Best and Lavrov defined generalised composition operations [10] for HLPN, which allow their compositional construction. The base operations of sequential composition, iteration, choice and refinement are defined in such a way that they tolerate liberal place types in HLPN. The authors introduced weak and strong versions of the composition with respect to manipulation with place types in the process of composition. They showed that associativity, commutativity and coherence with respect to unfolding are preserved in the composition process.

*Hierarchical decomposition*

In [11] Felder, Ghezzi and Pezze introduce the mechanism of formal hierarchical decomposition of timed HLPN (HLTPN). They provide the conditions for preserving properties from an abstract level to a lower (refined) level and, in order to ensure this, they define the concept of correct refinement. They also give a set of constructive rules that can be applied to the net refinement so that the resulting net is the correct refinement.

*Petri net process de/composition*

Tiplea and Desel [12] took another viewpoint. In their approach a Petri net is a set of processes, represented by the given net. Accordingly, they introduced a concept of process sample of a net with respect to a subnet and the so-called jumping nets. They showed how any process that is a composition of two nets can be decomposed in processes of component nets and vice versa. They use jumping nets as a tool for process sample generation and show how the samples can be generated. Jumping nets are, informally put, subnets of a net with a defined set of connecting places to connect them and a set of jumps from one subnet to another. The jumps are

relations between the markings of subnets and each jump is executed when a token reaches a connecting place and the relevant enabled transition is fired, so the token “jumps” from one subnet to another.

### Algebraic approach

In general, algebra offers an elegant way to compose objects (e.g. processes in process algebras). Moreover, close relationships exist between PN and algebras (e.g. algebraic PN). This has inspired some researchers to use algebraic methods for de/composition.

In [13] Rojas considers the so-called compositional well formed nets (cWN) with a set of defined composition operators. The composition operators are based on ideas of those from process algebras. Rojas studies the construction of structural and state space information from components and defines the matrix analysis of cWN from its components. It shows how to analyse cWN and the concept is extended to stochastic cWN with a time aspect and proposed methods for compositional behaviour.

Petri net composition with trace theory is proposed in [14]. Trace theory is a formalism that describes systems by their traces (sequences of symbols). In this work, by taking the principles of trace theory Coomber provides the composition technique, which forms a single Petri net from two Petri nets with joint behaviour of these nets.

### B. Related problems

#### Redundant information analysis

In [15] Da Siviera, Combacau and Portela address the problem of redundant information analysis in a decomposed Petri net model. They noticed that after using certain decomposition techniques the decomposed model may contain some redundant information. Such information may, in many cases, be relevant because of dependencies between modules from the duplication of critical information. They proposed a set of indicators to help designers make decisions and to evaluate the consequences of introducing redundant information. In real life systems, redundancies increase costs and reduce the performance of systems.

#### Avoiding state explosion problem

The state explosion problem (SEP) is closely related to the state space of PN (state space is closely related to RP and other analytical problems). Basically, SEP is a problem of enormous or infinite set of reachable states (even for small PN), which is hard to analyse. Many works are devoted to SEP for different classes of PN. Petrucci and Finkel [16] studied de/composition based on a common subset of places or transitions and provided algorithms to calculate the minimal coverability graph of a whole net by composing minimal coverability graphs of its subnets.

#### (Generalized) Stochastic PN

Due to the several different characteristics of (Generalized) Stochastic PN ((G)SPN) which consider and model stochastic processes (using the theory of probability), slightly different approaches are used in the de/compositional analysis with respect to timing aspects of this class of PN.

Li and Woodside focused on SEP and computational complexity for SPN in [17] and proposed a reduction technique and an iterative delay equivalent reduction

technique (IDERT) to solve the reduced SPN. IDERT iteratively tunes partly aggregated auxiliary SPNs until there is a delay approximately equivalent to the original SPN. Such an approach leads to an approximation of a more general SPN.

Campos and Perez-Jimenes [18] dealt with net decomposition techniques in order to reduce the SEP for the computation of performance indices for SPN. The basic idea is to (partially) represent the reachability graph of a given SPN in a decomposed manner, which can be used for exact or approximated performance analysis.

### C. Special de/compositional classes

Instead of developing methods for the de/composition of existing classes, some authors proposed new ones especially designed to take full advantage of the de/compositional approach. These classes are, basically, extensions of existing classes and their analysis is extended with respect to de/compositional characteristics. In the following, some of them are presented briefly.

#### Modular analysis and M-nets

The main goal of modular analysis introduced by Petrucci [19] is to cope with the SEP by reducing the state space. In order to achieve this goal there was a new class of PN introduced, the so called M-nets (modular nets), which allows the design of systems in a modular way. Very useful properties of modular design are: reusable modules, model structuring and functional de/composition. Modules imply modular state space (set of local state spaces of particular modules) easier to analyze, composition of subsystems invariants and modular, compositional or incremental verification (model refinement, state space construction...). The proposed construction algorithm and algorithm for constructing synchronization graph between modules allow taking advantage of the modular design.

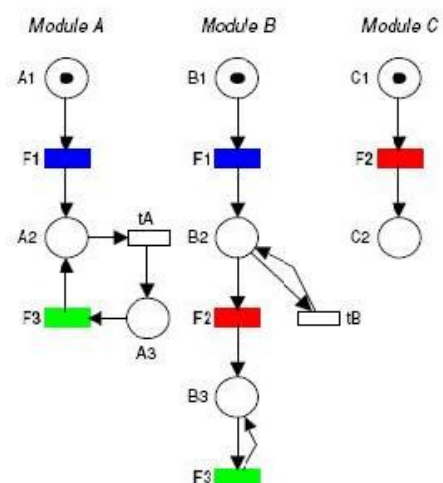


Fig. 3. Modular system design

The model refinement allows the design from abstract to lower levels with three possible types of refinement: type refinement (each value of refined type can be projected onto a value of the abstract type), subnet refinement (augmenting a subnet with additional places, transitions and arcs) and node refinement (replacing a place or a transition by a subnet). There are also proposed algorithms for investigating standard properties in the modular design, such as deadlocks,

reachability and liveness. The achieved results are extended to time and timed PN with respect to their time characteristics.

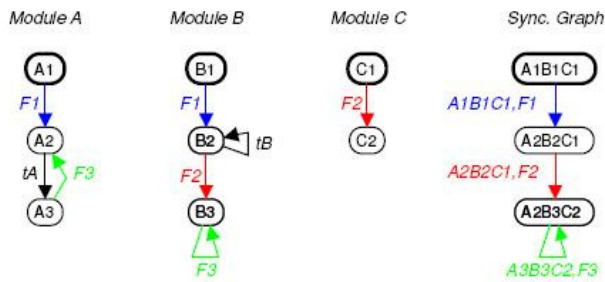


Fig. 4. Modular state space and synchronisation graph

#### Compositional Timed PN

Wang, Deng and Zhou [20] introduced a special compositional time Petri net (CTPN) extending time Petri Net (TPN). These extensions are basically “modularised” TPN that consist of components and connectors used to connect components. They also proposed reduction rules for TPN (component-level reduction rules), which transform a TPN component to very simple one by preserving the external observable timing properties. As a result, the proposed method makes it possible to reduce the size of TPN and, as the proposed method works at a coarse level (rather than an individual transition level), significantly fewer applications are required to reduce the size of the TPN compared to existing ones.

#### Semi-nets

In [21] Winkowski defines a generalised concept of PN with inputs absorbing information and outputs emitting information. These structures are called semi-nets and there is a concept of behaviour similar to PN behaviour introduced in the paper. The author defines partial composition operations for semi-nets by combining inputs and outputs so the resulting semi-net represents the joint behaviour of its subnets.

## IV. CONCLUSION

This paper considered some of the existing approaches developed for the de/compositional analysis of PN. Analytical problems, techniques and de/compositional approaches were presented briefly. The selection of the presented approaches is not adventitious – the aim is to point to general ones, instead of those that are problem-oriented (many works are devoted to de/composition in order to solve a concrete problem, e.g. the composition of web services based on PN, task decomposition in robotic assembling based on PN and so on).

In general, it can be seen that the presented approaches have some principles in common: resulting subnets after decomposition have some common places and/or transitions (they were duplicated) and composition merges relevant places and/or transitions. Moreover (not surprisingly), de/composition affects PN analysis in the sense of modularity by defining appropriate rules for it, and it allows the modular analysis of “sufficiently” large PN and avoiding SEP.

## REFERENCES

- [1] T. Murata, “Petri nets: properties, analysis and applications“, *Proceedings of the IEEE*, 1989, <http://wpage.unina.it/cotroneo/dwnd/Vittorini/Murata.pdf>
- [2] Š. Hudák, *Reachability analysis of systems based on Petri nets*, Košice, TU, 1999, ISBN: 80-88964-07-5, ch. 6
- [3] H. C. Yen, “Introduction to Petri net theory”, Dept. of Electrical Engineering, National Taiwan University, Taipei, Taiwan, <http://cc.ee.ntu.edu.tw/~yen/papers/yen.pdf>
- [4] D. A. Zaitsev, “Functional Petri nets”, *Universite Paris-Dauphine, Cahier du Lamsade* 224, Apr, 2005, pp. 62, <http://www.lamsade.dauphine.fr/cahiers.htm>
- [5] D. A. Zaitsev, “Decomposition of Petri nets”, *Cybernetics and Systems Analysis*, no. 5, 2004, pp. 131-140
- [6] D. A. Zaitsev, “Decomposition-based calculation of Petri net invariants”, *Proceedings of Workshop on Token based computing (ToBaCo), Satellite Event of the 25-th International conference on application and theory of Petri nets*, Bologna, Italy, June 21-25, 2004 / Cortadella, Yakovlev (Eds.), June, 2004, pp: 79-83
- [7] Q. Zeng, X. Hu, J. Zhu, H. Duan, “A polynomial-time decomposition algorithm for a Petri net based on indexes of places”, *Journal of Applied Sciences* 8 (24): 4668-4673, 2008, ISSN 1812-5654, 2008, Asian Network for Scientific Information
- [8] Q. Zeng, “Language, liveness and fairness invariant in decomposition of Petri Net based on the index of place”, <http://www.dcc.ufla.br/infocomp/artigos/v5.1/art02.pdf>
- [9] Q. Zeng, “Behavior descriptions of structure-complex Petri nets based on synchronous composition”, *Journal of Software*, 2004,15(3): pp. 327-337, <http://www.jos.org.cn/1000-9825/15/327.htm>
- [10] E. Best, A. Lavrov, “Generalized composition operations on high level Petri nets”, *Fundamenta Informaticae* 34, IOS Press, 2000, pp. 1-39, <http://parsys.informatik.uni-oldenburg.de/~best/publications/best-lavrov-composop.ps>
- [11] M. Felder, C. Ghezzi, M. Pezze, “Hierarchical decomposition of high level timed Petri nets”, *IPTEs Report, IPTES-PDM-54-V2.0*, <ftp://ftp.ifad.dk/pub/iptes/letter/iptes-pdm-54.ps.gz>
- [12] F.L. Tiplea, J. Desel, “Petri net process decomposition with application to validation”, <http://SunSITE.Informatik.RWTH-Aachen.DE/Publications/CEUR-WS/Vol-26/desel.pdf>
- [13] C. M. Isabel Rojas, “Compositional construction and analysis of Petri net systems”, *Dissertation, University of Edinburgh*, 1997, <http://www.era.lib.ed.ac.uk/handle/1842/421>
- [14] C.J. Coomber, “Petri net composition with trace theory”, <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.25.8059>
- [15] M. Da Silveira, M. Combacau, E. Portela, “Redundant information analysis in a decomposed Petri net model”, *Computational Intelligence for Modelling, Control and Automation*, 2005 and *International Conference on Intelligent Agents, Web Technologies and Internet Commerce, International Conference on Volume 2*, Issue , 28-30 Nov. 2005, pp. 753 – 758
- [16] A. Finkel, L. Petrucci, “Avoiding state explosion by composition of minimal covering graphs”, *Rapport de Recherche CEDRIC 91*, Centre d’Etudes et de Recherche en Informatique du Conservatoire National des Arts et Metiers, Paris, France, 1991
- [17] Y. Li, C. M. Woodside, “Iterative decomposition and aggregation of stochastic marked graph Petri nets”, *Proceedings of the 12th International Conference on Application and Theory of Petri Nets*, 1991, Gjern, Denmark, Jun, 1991, pp. 257-275
- [18] C.J. Perez-Jimenez, J. Campos, “On state space decomposition for the numerical analysis of stochastic Petri nets”, *Proc. 8th Int. Workshop on Petri Net and Performance Models (PNPM’99)*, 8-10 October 1999, Zaragoza, Spain, 1999, pp. 32-41
- [19] L. Petrucci, “Modularity and Petri nets”, *Proc. 7th Int. Symposium on Programming and Systems (ISPS’2005)*, Algiers, Algeria, pp. 7-8, May 2005, <http://hal.archives-ouvertes.fr/docs/00/04/07/22/PDF/LP-ISPS05.pdf>
- [20] J. Wang, Y. Deng, M. Zhou, “Compositional time Petri nets and reduction rules”, *IEEE Transaction on Systems, Man and Cybernetics*, Vol. 30, Part B, No. 4, Aug, 2000, pp. 562-572
- [21] J. Winkowski, “A Generalization of Petri nets by equipping them with inputs and outputs”, *ICS PAS Report 686*, Warsaw, Poland, Polish Academy of Sciences, Institute of Computer Science, 1990, pp. 1-18

# Context for concepts

Viliam ROČKAI

Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

viliam.rockai@gmail.com

**Abstract**—Development of an algorithm which would assign a name to a set of tokens, without supervisor, appears to be an unsolvable problem. While it’s easy for us to classify words “red, green, blue” as “colors” or “cat, dog, horse” as “animals”, computers need to be supported by a set of rules or hierarchies written by a man. In this paper, a method used to denote such token sets without supervisor is presented. The method uses associative learning of concepts to compute context strength between a group of tokens and a potential name for the group.

**Keywords**—associative learning of concepts, confabulation theory, natural language processing.

## I. INTRODUCTION

This paper presents the state of the art in labelling a set of tokens with another token, using knowledge derived from associative learning of concepts. It is a step forward in creation of a system for conversion of thoughts written down in a natural language into conceptual graph models [1]. When analysing such a thought (which can be a sentence or a phrase), the system needs to understand each word in a wider context. It needs to be able to find other words within the context and it needs to know which word to use to label the context. For example in a model sentence “the apple was red and fresh” the system should be able to understand the word “red” as an element of a set “yellow, green, blue, red, brown, ...”, which can be labelled with the name “color”. In [2], a method for computing token relatedness using associative learning of concepts has been presented. In this paper, we present a method extending our previous work in the field. The theory [2] is based on a neuro-physiological model of thalamic-cortical information processing by R. Hecht-Nielsen [3] which was newly published and summarised in [4]. Associative learning of concepts works with a fixed lexicon of symbols, which are implemented as strings (words of the natural language in question) in the current work. It is an unsupervised learning process carried out on a stream of symbols. The dynamics of this learning are modelled by Hebbian learning method. The learning process consists of introducing so-called associations between these symbols. Weighted associations stored in matrices called fascicles are the fundamental element of the associative learning of concepts.

## II. ASSOCIATIVE LEARNING OF CONCEPTS

The learning process consist of storing co-occurrences of a token (word) pairs in four contextual distances. Contextual distance is meant to be the number of words between that token pair within the sentence. The learning is done over a stream of symbols (i.e. English literature). It consists of two main phases. In the first phase, a lexicon of fixed number of symbols (words) is generated. Then, in the second phase,

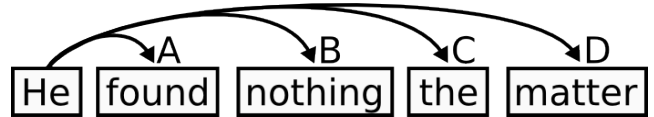


Fig. 1. Learning process is being done over a stream of tokens. In this case, co-occurrences are stored in 4 separate matrices (A, B, C, D), called fascicles.

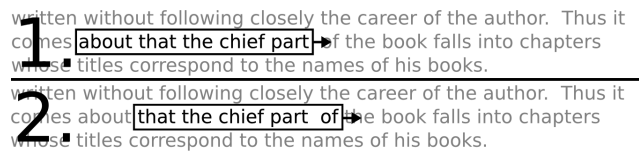


Fig. 2. Token window is moving forward through the text stream. The knowledge is stored into fascicles only if all its tokens are present in the lexicon.

associations between those symbols are inducted in a number of contextual distances (fig. 1).

The storage of these word co-occurrences uses a token window. This token window shifts forward through the token stream and, in the case, when all the tokens in the window are known (can be found in lexicon), the co-occurrences of all contextual distances are stored (fig. 2).

These co-occurrences are then used for computing the probabilities used below. For deeper understanding, more information about the learning process can be found in [2].

Establishment of an association between two symbols depends on significance of the co-occurrences. The formula for determining this significance within a contextual distance  $d$  for such an association used in this paper is:

$$sig_d(i, j) = p_{i,j} \log_2 \left( \frac{p_{i,j}}{p_i p_j} \right) \quad (1)$$

It is derived from the formula for measurement of mutual information content of two discrete random variables [5]. It expresses mutual dependence of two variables ( $i$  and  $j$ ).  $p_i$  and  $p_j$  are probabilities of seeing words  $i$  or  $j$  in the text.  $p_{i,j}$  then stands for the probability of seeing these two words together. When the significance of some token pair is greater than 0, an association between them is established. The weight of association is irrelevant for context strength computing: only the information about existence of such an association is important to us.

## III. CONTEXT COMPUTING

Ambiguity is one of the crucial problems in natural language processing. Words have different meanings in different environments. A “mouse” can stand for “mammal” or “computer device”. In some situations it is more suitable to label it as a

“pet”, “rodent” or even as “food”. In this paper, it is assumed that the “meaning” of such a word can be derived from its environment. Association between the word and its meaning should exist in some contextual distance. Our paper gives an implementation of this assumption. A method is used to find and quantify all available contexts for a set of words. In our experiments, context is represented as a one word phrase.

To be able to filter relevant contexts from the rest, we need to quantify the value of the contextual strength between a set of tokens and the context token. Context strength for a token pair is defined as a weighted sum of significances of those tokens through all contextual distances with the relevant formula divided by probability of seeing the context word in the input text:

$$c(x, y) = \sum_{d=0}^n \frac{w_d \text{sig}_d(x, y)}{p_y} \quad (2)$$

Here,  $x$  stands for the input word and  $y$  stands for its context. The strength is computed for  $n$  contextual distances. Dividing with value of  $p_y$  is used to filter out stop word tokens which are often associated with many words. The significances for input word and context word are computed only in the case when these two are associated in given contextual distance. Finally, the context strength for a set of tokens  $T$  and context token  $x$  is defined below:

$$c(T, y) = \min(c(t_0, y), c(t_1, y), \dots, c(t_n, y)) \quad (3)$$

The strength of context  $y$  for a token set  $T$  is given as the minimum value of contextual strengths among all words from the set.

#### IV. EXPERIMENTS

Computation of context was tested on a set of word triads. For each triad, the anticipated context was guessed, and then the context strength was computed for the triad and each word in the lexicon. Context candidates were then ordered by their strength. The results are presented in Table 2. Software used for this experiment can be obtained (with Java source code) from [http://neuron.fei.tuke.sk/~rockai].

The system has been working with a lexicon of 9000 words. The words have been chosen as the 9000 most frequent words from six random English books obtained from the Gutenberg library project [http://www.gutenberg.org] (i.e. The Life and Letters of Charles Darwin, Volume I; Thomas Wingfold, Curate; ...). The associations were accumulated from a bigger corpus and also obtained from the Gutenberg project. The corpus consisted of 1 GB of randomly chosen English literature in plain text format. Example text from the corpus can be seen at the end of this section, in Subsection A.

The associations have been accumulated using a 5-token window, into 4 separate fascicles (for 4 contextual distances). The knowledge has been stored only in the case where all tokens in the window were present in the lexicon. The weights used for computing contextual strength are shown in Table 1.

Context has been computed for chosen word triads. The result of the operation is a list of all tokens from the lexicon, ordered by context strength for current triad. For each word triad, one anticipated context had been guessed. The guess had been done by a human, based on practical world knowledge.

TABLE I  
WEIGHT USED FOR CONTEXT

Contextual distance (d)	Weight ( $w_d$ )
0	1
1	0.5
2	0.25
3	0.125

TABLE II  
DETERMINING OF CONTEXT WORD FOR TOKEN TRIADS

Gussed context	Word triad	Rank	Winner
dish	chicken,potatoes,fish	3	roast
art	painting,music,poetry	4	Italian
drink	beer,wine,water	4	bottles
plant	tree,flower,wheat	5	grows
literature	prose,roman,poetry	8	poets
instrument	piano,guitar,organ	8	plays
war	fight,battle,death	8	fight
waters	lake,river,sea	10	frozen
cloth	skirt,trousers,coat	10	silk
weapon	bow,gun,sword	15	swung
color	red,green,blue	19	velvet
royal	king,queen,prince	39	palace
building	house,palace,church	46	build
fruit	orange,peach,cherry	50	blossoms

For each triad, a context word has been guessed. The rank column shows the position of the guessed context word in the ordered context candidates list. Winner context column shows the word which appeared first in the context candidate list.

The resulting list can be seen Table 2. “Rank” stands for the position of guessed context among all candidate context lists, ordered by their strengths. The winner column stands for the first word from that list. Since the context strength is a very domain specific variable, the winner can be understood as the most probable context for input words within the domain of input documents used in the learning process.

Word triads show some promising results. The more information the system has about the word cluster, the more easily it can choose better candidate to label such a set. Adding an additional word to each triad has been assumed to be a good refinement method to achieve better results in context candidate rank value. Table 3 consists of queries made upon word 4-tuples, where additional word has been chosen for each set in the meaning of guessed context. In rank column the progress of guessed context is written in parentheses (when compared to previous table). Only the set of words that should fall into “waters” context has shown regression.

##### A. Corpus sample

... “Here! What do you want? Go away from that shop! No one is allowed there!” and looking from an upper window, Tom saw his father running toward a stranger, who was just stepping inside the shop where Mr. Swift was constructing his turbine motor. Tom started as he saw that the stranger was the same black-mustached man whom he had noticed in the post-office, and, later, in the restaurant at Mansburg...



TABLE III  
DETERMINING OF CONTEXT WORD FOR TOKEN TRIADS

Guessed context	Word foursome	Rank	Winner
dish	chicken,potatoes,fish,pie	1 (2)	dish
art	painting,music,poetry,literature	2 (2)	Italian
drink	beer,wine,water,tea	2 (2)	drank
plant	tree,flower,wheat,corn	4 (1)	grows
literature	prose,roman,poetry,writings	2 (6)	poets
instrument	piano,guitar,organ,harp	6 (2)	plays
war	fight,battle,death,enemy	6 (2)	fight
waters	lake,river,sea,pond	17 (-7)	margin
cloth	skirt,trousers,coat,jacket	8 (2)	silks
weapon	bow,gun,sword,knife	8 (7)	sword
color	red,green,blue,yellow	8 (11)	silks
royal	king,queen,prince,princess	19 (20)	palace
building	house,palace,church,hospital	24 (22)	build
fruit	orange,peach,cherry,apple	44 (6)	blossoms

There was one additional word added to every triad from the previous table. The words were chosen to refine the meaning and therefore strengthen the context strength. In rank column the progress of guessed context is written in parenthesis (when compared to the previous table).

## V. CONCLUSION

Automatic identification of a name for some set of words is a tricky problem in computer science. Computers need to deal with vagueness and domain dependency of natural language. Associative learning of concepts can contribute to this area of research. Thanks to computation of context strength, the system can suggest a set of words for candidates for such a naming. More work and study needs to be done in context strength computing, but examples shown in Table 2 indicate very promising results. The next research should be focussed on corpus preprocessing. Since naming of word clusters is a very domain-specific task, experiments should be done over domain-specific corpuses. One of the goals of continuing research should be a more objective way to measure the results.

## ACKNOWLEDGMENT

The work presented in the paper was supported by the Slovak Grant Agency of Ministry of Education and Academy of Science of the Slovak Republic within the 1/4074/07 project "Methods for annotation, search, creation, and accessing knowledge employing metadata for semantic description of knowledge".

## REFERENCES

- [1] J. F. Sowa, *Conceptual Structures: Information Processing in Mind and Machine*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 1984.
- [2] V. Ročková, *Mining of Concepts and Semantic Relations from Texts in Natural Language*. Košice: Technická univerzita, 2005.
- [3] R. H. Nielsen, "A Theory of the Cerebral Cortex," *5th Int Conf Neural Inf Proc Vol 3*, vol. 3, pp. 1459–1464, 1998.
- [4] R. H. Nielsen, *Confabulation Theory: The Mechanism of Thought*. Springer, 2007.
- [5] T. M. Cover, *Elements of Information Theory*. Wiley-Interscience, July 2006.

# Survey on Support for Design Patterns in Software Application Development

*Miroslav SABO*

Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

miroslav.sabo@tuke.sk

**Abstract**—Design patterns are general reusable solutions to recurring problems and nowadays widely used in process of software development. They have proven to be very useful for the design of object-oriented systems as they elevate the level of abstractions by capturing a relationship among the language level abstraction mechanisms. The implementation of the higher abstractions, however, suffers from a number of problems due to the fact that insufficient language support is provided by the traditional object-oriented paradigm. This paper addresses this problem and presents the results of survey conducted in the area of language and tool support for the utilization of design patterns in software application development.

**Keywords**—design patterns, language support, software development, tool support.

## I. INTRODUCTION

A high-level programming language is a programming language with strong abstraction mechanisms which should be easy to use, highly understandable and widely portable across platforms. Nonetheless, the abstraction mechanisms of current object-oriented programming languages are still too low-level. Design patterns elevate the level of available object-oriented abstractions by capturing a relationship among the language-level abstraction mechanisms. They have proven to be very useful for the design of object-oriented systems. The power of Design Patterns stems from their ability to provide generic solutions to reappearing problems that can be specialized for particular situations. The implementation of the Design Patterns, however, suffers from a number of problems due to the fact that insufficient language support is provided by the traditional object-oriented paradigm.

## II. PROBLEMS AND CHALLENGES IDENTIFIED WHEN USING DESIGN PATTERNS

Design Patterns are techniques that push the advantages of the object-oriented paradigm even further by providing, among others, additional reusability of both design and implementation specifications. Despite these advantages, one can identify some problems or obstacles related to the implementation of Design Patterns. In [3], it is stated, that these problems are caused by the fact that object-oriented languages are rigid, whereas the object-oriented paradigm constantly develops and regularly is extended with new concepts. These problems are primarily related to the traceability of design patterns in the implementation, the self

problem, language expressiveness and the implementation overhead of design patterns:

- *Traceability*: The traceability of a Design Pattern is often lost because the programming language does not support a corresponding concept. The software engineer is thus required to implement the pattern as distributed methods and message exchanges, i.e. the pattern which is a conceptual entity at the design level is scattered over different parts of an object or even multiple objects. This problem has also been identified by [19].
- *Reusability*: Design Patterns are primarily presented as design structures. Since design patterns often cover several parts of an object, or even multiple objects, patterns have no first class representation at the implementation level. The implementation of a design pattern can therefore not be reused and, although its design is reused, the software engineer is forced to implement the pattern over and over again.
- *Implementation Overhead*: The implementation overhead problem is due to the fact that the software engineer, when implementing a Design Pattern, often has to implement several methods with only trivial behavior, e.g. forwarding a message to another object or method. This leads to significant overhead for the software engineer and decreased understandability of the resulting code.

Numerous authors have identified another challenges that arise when patterns are concretized in a particular software system. The three most important challenges are related to implementation, documentation, and composition [10].

Design pattern implementation usually has a number of undesirable related effects. Because patterns influence the system structure and their implementations are influenced by it, pattern implementations are often tailored to the instance of use. This can lead to them “disappearing into the code” [7] and losing their modularity [19]. This makes it hard to distinguish between the pattern, the concrete instance and the object model involved. Adding or removing a pattern to/from a system is often an invasive, difficult to reverse change [5]. Consequently, while the design pattern is reusable, its implementations usually are not [19].

The invasive nature of pattern code, and its scattering and tangling with other code creates documentation problems [19]. If multiple patterns are used in a system, it can become difficult to trace particular instances of a design pattern,

especially if classes are involved in more than one pattern (i.e. if there is pattern overlay/composition) [1].

Pattern composition causes more than just documentation problems. It is inherently difficult to reason about systems with multiple patterns involving the same classes, because the composition creates large clusters of mutually dependent classes. This is an important topic because some design patterns explicitly use others patterns in their solution.

### III. CATEGORIZING DESIGN PATTERNS

In [8] the authors state that one person's Design Pattern can be another person's primitive building block, because the point of view affects one's interpretation of what is and what is not a Design Pattern. And the point of view is influenced by the choice of programming language. It is said: "The choice of programming language is important, because it influences one's point of view. Our patterns assume SMALLTALK/C level language features, and that choice determines what can and cannot be implemented easily. If we assumed procedural languages we might have included design patterns called "Inheritance", "Encapsulation", and "Polymorphism". Similarly, some of our patterns are supported directly by less common object-oriented languages." Thus, they believe that Design Patterns do not need to be language independent.

#### A. Restrictive Analysis of Design Patterns

Agerbo and Cornils present in their paper [1] an analysis in the form of a set of criteria, that they have used for an evaluation of the Design Patterns that are presented in [8]. The analysis does not go so far as to identify the true Design Patterns and throw away the rest; instead it focuses on assembling a core of Fundamental Design Patterns (FDP) which should capture good Object Oriented design on a high enough level so that it can be used in various kinds of applications. The Design Patterns that are not judged to be Fundamental are then either classified differently or rejected completely. In [1] they have shown that by using the guidelines of this analysis, they can remove half of the Design Patterns from the core of Fundamental Design Patterns, so that out of the original 23 Design Patterns only 12 remain.

If concentrating on building a core of Fundamental Design Patterns that are not covered by any generally accepted language construct, one can use this core to form the common vocabulary to be used among computer scientists regardless of background. However, a Design Pattern which is covered by a language construct in one language might still be a design idea worth preserving in languages which does not have this language construct. Therefore the Design Patterns, which are not Fundamental because they are language dependent must be kept as Language Dependant Design Patterns (LDDPs).

For each of the Design Patterns in [8], Agerbo have discussed whether it is covered by a known object oriented language construct (and thereby an LDDP), an application of another Design Pattern (an RDP) or an inherent way of thinking in object-oriented programming. By partitioning the Design Patterns into:

- Fundamental Design Patterns,
- Language Dependant Design Patterns and
- Related Design Patterns,

Agerbo and Cornils have a core of the Design Patterns - the Fundamental Design Patterns - which fully provides the

benefits of Design Patterns. Only 12 of the 23 Design Patterns from [8] are classified as Fundamental Design Patterns following these criteria. This leads to conclude that it is beneficial to have a critical approach to Design Patterns, because it minimizes the amount of Fundamental Design Patterns and thereby makes the area of Design Patterns easier to get on top of.

#### B. Taxonomy of the Design Patterns

The main hypothesis of Gil and Lorenz is that in many ways, Design Patterns are nothing but "puppy language features" [9]. That is, patterns could (sometimes even should) grow to be language features, although they are not yet implemented as such. One evidence for this hypothesis comes from considering language-independence. Many design and analysis methodologies and notations pride themselves for being "language independent". Design patterns candidly admit they are not. "Point of view affects one's interpretation of what is and isn't a pattern. One person's pattern can be another person's primitive building block" and "the choice of programming language is important because it influences one's point of view". In essence, they adopt this observation and take it a step further, postulating that many patterns can be found as a feature of some language, but not the one in which the pattern is being applied.

They present taxonomy of patterns based on how far they are from becoming actual language features. In this taxonomy, low-level patterns are mere *clichés*. *Idioms* are patterns mimicking features found in another language. More concretely, idioms sometimes cater for the absence of these features in the current language. *Cadets* are patterns which are candidates for making their way into a programming language. They distinguish between two kinds of cadets. *Relators* are those patterns which capture the relation between a small number (typically two) of lingual entities such as objects, classes, etc. *Architects*, on the other hand, are patterns which describe the architectural structure of a large number of entities. Architects grow out of their lingual infancy into new language paradigms.

#### C. Alternative Pattern Representations

A number of papers address problems with the preciseness of the pattern description format presented in [8]. In [13] a hierarchical model (consisting of three layers based on UML notations) is introduced for describing pattern structures and dynamic behavior. The role model captures the "pure pattern", and is refined by a type-model (similar to the GoF UML diagrams), which is in turn refined by an instance-specific model that uses the concrete names a particular pattern instance. The authors claim that the three models complement each other and that a developer should have access to all three models of a particular pattern.

Florijn suggests a fragment-based representation of design patterns [7]. A fragment depicts a design element such as a class, method or association). Patterns themselves and all elements in a pattern instance (classes, relationships among them, code) are represented as (graphs of) fragments.

Mapeldsen introduce the design pattern modeling language DPML, built upon similar concepts as UML [14]. This multi-level approach (design patterns, pattern instances, and object

models) makes it possible to show objects and their roles within the pattern.

Mikkonen addresses the problem that the temporal behavior of Design Patterns is difficult to reason about and proposes a formal notation for this purpose [15]. This model formalizes patterns as behavioral layers, and realizes the interactions between objects as atomic actions. With this approach, pattern compositions can be modeled.

#### IV. APPROACHES TO IMPLEMENTING DESIGN PATTERNS

Initially design patterns were proposed primarily as design concepts, but during recent years the importance of corresponding language concepts has been recognized and several efforts to provide this can be identified. Providing language support for Design Patterns, however, is a difficult problem. The reason is that Design Patterns represent a concept that is orthogonal to the concepts provided by the programming language. Whereas a class provides a definition that covers the complete semantics of a single entity, design patterns, on the other hand, provide a description that covers part of the semantics of a collection of entities. When providing language support for design patterns, this requires that, for each class that plays a role in the design pattern, the partial specification by the design pattern is composed with the existing behavior of the class. Part of this composition of behavior may be done through multiple inheritance or aggregation, but there are situations where the pattern specific behavior needs to be superimposed on the class behavior. The latter type of composition is not supported in traditional object-oriented languages.

Due to these difficulties, several different approaches to providing this support have been developed. These approaches can be divided into the following categories:

- *Design environment support*: The simplest level of support is no form of language support but rather support through the design environment. During design the software engineer makes use of design patterns that have a representation in the environment. After implementation, the environment is able to make the code related to a design pattern visible in the code. Since a design pattern often incorporates multiple classes, the visualization should also incorporate this. The good example of such an approach, as well as the following one, is Design Pattern Automation Toolkit [6].
- *Generative approach*: An alternative approach is to generate code from a design model that comprises Design Patterns. For each pattern used in the design, code is generated in the classes that make up the design. A class may be involved in multiple Design Patterns, causing the generated code for the class to be composed from these patterns. After code generation, the application-specific code for the class can be added to the class specification. An example of this approach is presented in [5].
- *Design pattern libraries*: Since design pattern descriptions contain information about how the participants interact with each other, what interfaces and variables they have to have, it is only natural to investigate how much of the design and code generation process can be automated. In [19] it is proposed to build the library of Design Patterns

consisting of so-called pattern classes. The most obvious way of using a library of Design Patterns is by letting the classes in the application inherit from the classes in the LDP. A pattern class encapsulates all the behavior and logic of the Design Pattern and the classes that form the Design Pattern in the application thus contain no methods related to the Design Pattern. What are left in the classes are only pointers and other data required for the Design Pattern. The problem of this solution is that all the structure of the Design Pattern is lost, since everything is now contained as methods in the pattern class. How and to what extent the fundamental Design Patterns could be placed in a library of Design Patterns is also discussed in [1].

- *Programming language extensions*: The most complete approach is when the programming language itself provides direct support for representation of design patterns. The difficulty with this approach is that the composition of design pattern behavior and class behavior needs to be supported, including those situations where the design pattern behavior needs to be superimposed on the class behavior. An example of the latter is the Observer pattern where the change notification behavior needs to be superimposed on the relevant methods of the class. An example of this approach can be found in [4].

##### A. Design Pattern Implementation in Java and AspectJ

Hannemann and Kiczales explore the effect of aspect-oriented programming techniques on the implementation of the GoF patterns in [10]. They claim that their implementations show modularity improvements in 17 of 23 cases. These improvements are manifested in terms of better code locality, reusability, composability, and (un)pluggability. The degree of improvement in implementation modularity varies, with the greatest improvement coming when the pattern solution structure involves crosscutting of some form, including one object playing multiple roles, many objects playing one role, or an object playing roles in multiple pattern instances.

Nordberg describes how AOP and component-based development can help in software module dependency management [16]. In a different work, he views design pattern improvements from the point of view of indirections and shows how replacing or augmenting OO indirection with AOP indications can lead to better designs [17].

##### B. Design Pattern in Dynamic Programming

Norvig's work on design patterns in dynamic programming [18] explores impacts on the GoF design patterns when implemented in Lisp and/or Dylan. This work is another indicator that patterns depend on the language paradigm. Of the 23 patterns, he found that 16 either become either invisible or simpler due to first-class types, first-class functions, macros, method combination, multimethods, or modules. He classifies the levels of implementation that a pattern may have into three groups:

- *Invisible*: So much a part of language that you don't notice (e.g. when class replaced all uses of `struct` in C++, no more "Encapsulated Class" pattern)
- *Informal*: Design pattern in prose; refer to by name, but must be implemented from scratch for each use
- *Formal*: Implement pattern itself within the language. Instantiate/call it for each use. Usually implemented with macros.

### C. Design Pattern through Reflection

Sullivan investigated the impact of a dynamic, higher-order OO language (Scheme with a library of functions and macros to provide object oriented facilities) on design pattern implementations [20]. In-line with Norvig's work, he observed that some design pattern implementations disappear (if language constructs capture them), some stay virtually unchanged and some become simpler or have different focus.

### D. Implementing Design Pattern with Layered Object Model

Bosch has studied the issue of language support for design patterns in his work on the layered object model (LayOM) [4]. The layered object model is an extended object model, i.e. it defines in addition to the traditional object model components, additional components such as layers, states and acquaintance categories. The layers encapsulate the object so that messages send to or by the object have to pass the layers. Each layer, when it intercepts a message, converts the message into a passive message object and evaluates the contents to determine the appropriate course of action. Layers can be used for various types of functionality. Layer classes have, among others, been defined for the representation of relations between classes and objects, design patterns and programming conventions.

### E. Language Support for Design Patterns using Attribute Extension

Hedin has in [11] presented a technique for formalizing design patterns using a technique based on attribute grammars. The technique allows Design Pattern applications to be identified in the source code, and supports automatic checking that the pattern is applied correctly. She states that although the pattern descriptions are based on semi-formal class diagrams which allow for many implementation variations, they are sufficiently precise so that when selecting a particular implementation, the pattern could be formalized and form the basis of some kind of programming language support.

She proposes the use of attribute extension [1] to support the specification and application of design patterns. It is a technique which allows the static-semantics of a language to be extended, allowing conventions to be enforced, but which keeps the syntax of the base language. A key advantage of this technique is that it is easy to integrate with existing languages and environments.

## V. CONCLUSION

In this paper I have presented the general overview of the area of design patterns in software development. The main problems of using patterns were listed in the beginning and various approaches to solving them were described in the

following sections. This theoretical research will serve as a basis for my future work, in which I will focus on the support for design patterns during application evolution, more specifically – problem of identification of pattern constituents in source code and checking that the design pattern is applied correctly.

## ACKNOWLEDGMENT

This work was supported by VEGA Grant No. 1/4073/07 Aspect-oriented Evolution of Complex Software Systems.

## REFERENCES

- [1] E. Agerbo, A. Cornils, "How to preserve the benefits of Design Patterns", in Proceedings of OOPSLA'98, pp. 134–143, 1998.
- [2] S.R. Alpert, K. Brown, B. Woolf, "The Design Patterns Smalltalk Companion", Addison-Wesley, 1998.
- [3] J. Bosch, "Design Patterns as Language Constructs", in the Journal of Object-Oriented Frameworks, Volume 11, Issue 2, SIGS Publications, 1998.
- [4] J. Bosch, "Design Patterns & Frameworks: On the Issue of Language Support", in Proceedings of ECOOP'97, 1997.
- [5] F. Budinsky, M. Finnie, P. Yu, J. Vlissides, "Automatic code generation from Design Patterns", in IBM Systems Journal, Volume 35, Issue 2, pp. 151–171, 1996.
- [6] DPAToolkit, on Internet <http://www.dpakit.sf.net>, 1.3.2009.
- [7] G. Florijn, M. Meijers, P. van Winsen, "Tool support for object-oriented patterns", in Proceedings of ECOOP'97, 1997.
- [8] E. Gamma, R. Helm, R. Johnson, J. Vlissides, "Elements of Reusable Object-Oriented Software", Addison-Wesley, 1995.
- [9] J. Gil, D.H. Lorenz, "Design Patterns vs. Language Design", in Proceedings of ECOOP'97, 1997.
- [10] J. Hannemann, G. Kiczales, "Design Pattern Implementation in Java and AspectJ", in Proceedings of OOPSLA'02, pp. 161–173, 2002.
- [11] G. Hedin, "Language Support for Design Patterns using Attribute Extension", in Proceedings of ECOOP'97, pp. 137–140, 1997.
- [12] G. Hedin, "Attribute Extension – a Technique for Enforcing Programming Conventions", in Nordic Journal of Computing, Volume 4, pp. 93–122, 1997.
- [13] A. Lauder, S. Kent, "Precise Visual Specification of Design Patterns", in Proceedings of ECOOP'98, pp. 114–134, 1998.
- [14] D. Mapeldsen, J. Hoskings, J. Grundy, "Design Pattern Modeling and Instantiation using DPML", in Proceedings of TOOLS'02, pp. 3–11, 2002.
- [15] T. Mikkonen, "Formalizing Design Patterns", in Proceedings of ICSE'98, pp. 115–124, 1998.
- [16] M.E. Nordberg III, "Aspect-Oriented Dependency Inversion", in Proceedings of OOPSLA'01, 2001.
- [17] M.E. Nordberg III, "Aspect-Oriented Indirection – Beyond Object Oriented Design Patterns", in Proceedings of OOPSLA'01, 2001.
- [18] P. Norvig, "Design Patterns in Dynamic Programming", in Object World '96, 1996.
- [19] J. Soukup, "Implementing Patterns", in J.O. Coplien, D.C. Schmidt: PatternLanguages of Program Design, Addison-Wesley, pp. 395–412, 1995.
- [20] G.T. Sullivan, "Advanced Programming Features for Executable Design Patterns", Lab Memo, MIT Artificial Intelligence Laboratory, number AIM-2002-005, 2002.

# Digital Audio Watermarking in MPEG-1 Audio Layer III: A Survey

Ján STAŠ

Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

jan.stas@tuke.sk, ing.jan.stas@gmail.com

**Abstract**—In the last couple of years, music distribution over the Internet has become an increasingly simple way how to receive audio recordings of popular bands and singers. In this context digital audio watermarking can provide an effective protection against illegal distribution, reproduction or copying of this audio content and also enables the robust and inaudible transmission of additional right information within the audio data. Whereas most of the known algorithms operate in the uncompressed or linear domain, few are capable of embedding watermarks into compressed domain. This paper presents the main techniques for an insertion of watermarks into compressed audio signals. One of the most generally used formats for distributing the audio content via Internet are MPEG-1 audio layer III (MP3) formats. This paper covers the basic theory behind MP3 audio compression and several basic techniques for inserting the watermark before and during the encoding step and directly in compressed audio bitstream will be presented. At the end of this paper the possibility of another application of these watermarking algorithms in real-time distribution by database system via Internet will be designed.

**Keywords**—Embedding, MPEG-1 audio layer III, spread spectrum, watermark.

## I. INTRODUCTION

Dynamic development of digital technologies for manipulation and transmission of the multimedia content has brought the specific problems, which resulted from digital form. This allows people to easily make and spread illegal copies of digital data, which brings up the problem of counteracting the authors rights, copyright. One of the options how to avoid this is to use digital watermarking technologies. In recent years were developed and verified various types of methods for embedding watermarks into multimedia and assign several types of systems that implement digital watermarking technology.

In the event of audio content, several methods for inserting watermarks into uncompressed audio signals were proposed, which are based on the direct modification of signal in time domain. Among the all proposed in recent years can be for example mentioned the Echo Hiding algorithms [1][2], or earlier works based on the insertion into the Least Significant Bits [2][3]. Further mentioned ones like the Phase Coding [2][3] or Phase Modulation algorithms or the Spread Spectrum (SS) modulation [2][3] insert the watermark into uncompressed material in transformation domain (frequency domain). Everyone of these methods holds a number of advantages and disadvantages which in previous work were described.

This work is especially oriented on embedding watermarks just into the compressed audio content, specifically for MPEG-1 audio layer III formats. First off all, a short overview

about watermarking and its requirements will be mentioned. Consequently, the generation of watermark and watermarking in time domain by SS modulation will be presented. As follows a description about MPEG encoder and decoder will be noted. Later on we will present a main technique for embedding watermarks before the compressing, algorithm which integrated watermarking into compress encoder, algorithm based on the partially decoding-watermarking-reencoding and finally inserting into metadata will be shown. At the end of this paper a basic concept of the system based on the database for administration watermarked audio files by unique ID with authorized access will be presented.

## II. AUDIO WATERMARKING

Watermarking can be understood as an embedding of the additional information into multimedia content in order to its protection so that this embedded information satisfies the requirements for robustness and imperceptibility in the original data. Onto these systems which operate with watermarks additional requirements and properties are given. There are several of them which are the most important:

**Robustness** is a requirement for resistivity of the used method against the attacks of unauthorized persons, whose aim is to remove or damage inserted watermark, or make its detection impossible, or discredit the results of detection.

**Imperceptibility** is a requirement for perceptual transparency of the watermark inserted into the original data set, at which the insertion of watermark has not imposed degradation in quality of the watermarked audio signal, or has made perceivable artifacts. Imperceptibility uses primarily imperfection of human receptors.

**Capacity** or **data rate** in the case of digital audio watermarking algorithms express the number of distinguishable watermarks, or number of bits, which can be inserted into the original data by specific algorithm without decrease of perceptive quality of the original audio data.

**Computational complexity** of algorithm depends on its application. It expresses the computational severity and the time needed in these systems with watermarks to perform the algorithm for embedding or detection of the watermark.

However, all of these requirements together are not performable. With the appropriate manner it is possible to achieve a compromise between these requirements. It depends on the algorithm which is applied.

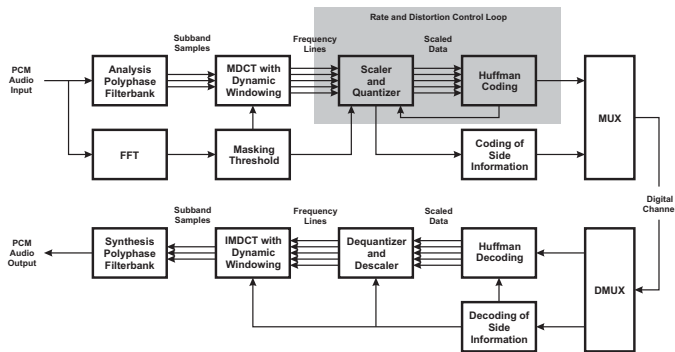


Fig. 1. Structure of MPEG-1 audio encoder and decoder layer III

### III. MPEG-1 AUDIO LAYER III

MPEG audio is a lossy algorithm, which uses the imperfection of human auditory system. It removes the perceptually irrelevant parts of the audio signal and makes the audio distortions inaudible to the human ear [4]. This standard is based on MUSICAM<sup>1</sup> and ASPEC<sup>2</sup> audio algorithms. MPEG audio compression standard supports three audio sampling rates at 32, 44.1, or 48 kHz. There are four different channel modes [5]:

- 1) Single mode - monophonic channel,
- 2) Dual mode - two independent monophonic channels,
- 3) Stereo mode - left and right channel,
- 4) Joint stereo mode - takes advantages of either the correlations between the stereo channel or irrelevancy of the difference between channels or both.

#### Encoding

The audio signal with length of 576 PCM samples (what is forming one of two granules of 1152 frame block) is introduced to the input of Polyphase Filterbank (see Fig. 1), which divides the audio signal into 32 equal width frequency subbands. Equal width filters do not correspond with critical band model of auditory system. High frequencies of critical bands cover several subbands of filter bank.

At the same time, input audio samples pass through MPEG psychoacoustic model that determines the ratio of the signal energy to the masking threshold for each subband. The model analyses the audio signal and computes amount of noise masking available as a function of frequency. The encoder then uses this information to decide how to use limited number of bits to represent the input audio signal [5]. MPEG-1 audio layer III has two psychoacoustic models, which use the Fourier transformation for this mapping. Before Fourier transformation Hann weighting is applied to the audio signal. The Psychoacoustic Model I uses 1024-point sample window and identifies tonal components based on the local peaks of the audio power spectrum and calculates a masking threshold for each subband in the Polyphase Filterbank. The Psychoacoustic Model II uses a 1024 sample window and two calculations per frame and computes tonality index as a function of frequency.

<sup>1</sup>Masking Pattern Adapted Universal Subband Integrated Coding and Multiplexing (1989) - a method for audio base-band coding; the bit rate of the coding process is reduced by using a psychocoustic model.

<sup>2</sup>Adaptive Spectral Perceptual Entropy Coding of high quality music signals (1991) - combines high coding efficiency with the flexibility needed to work in different applications, from 64 kbps per channel with FM quality to 128 kbps for CD quality.

This index gives a measure of whether the component is more tone-like and noise-like [5].

The output from Psychoacoustic Model and signal in form of 32 frequency subbands and 18 samples in each of subband is further processed by using Modified Discrete Cosine Transformation (MCDT). The MDCTs further subdivides the subband outputs in frequency to provide better spectral resolution. Layer III specifies two different MDCT block lengths: a long block of 18 samples or a short block of 6. The long block length allows greater frequency resolution for audio signal with stationary characteristics, while the short block length provides better time resolution for transients [5].

Consequently, block of 576 MDCT coefficients is non-uniform quantized and Huffman coded. The quantization procedure is composed of two loops, inner loop for bit rate control and outer loop for quantization distortion control. Inner loop is computed if the bit budget is larger than the bits available the step size is increased until the bit consumption is acceptable. Outer loop calculates the quantization distortion in every scale factor band and compares it with the allowed distortion obtained from the psychoacoustic analysis. If it is larger than the allowed distortion, the step size is decreased to reduce the artifact audibility [6].

In the Huffman entropy coding step, quantized coefficients are divided into 3 groups, a big values, a small values and zero regions. Layer III uses variable-length Huffman codes to encode the quantized samples. The big values (low frequencies) are coded as pairs, divided into 3 subregions and used different Huffman tables. The small values (-1, 0, or 1) are coded as quadruples and used the independent Huffman table. And zero values are coded by run-length coding (RLC) [6].

Finally the bit stream is formatted. Each frame contains main data, a 4 byte header and possibly CRC code and side information (scale factors and ancillary data). Scale factors in side information data carries specified important parameters for quantization in decoding step. Frame header holds the basic structural information about bitstream. The structure of header and formatting the bit stream is shown in Fig. 2.

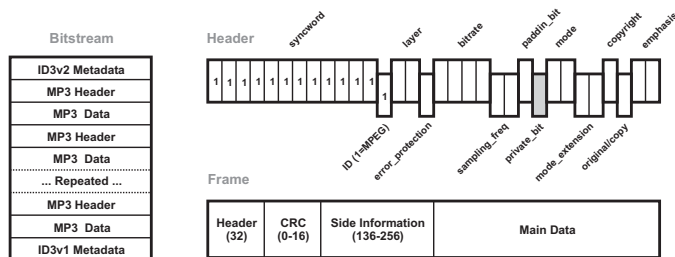


Fig. 2. MPEG format of header, frame and bitstream

#### Decoding

Decoding is the inverse process to the process of encoding (see Fig. 1). However, here is not used the Psychoacoustic Model. It is composed of bit stream decoding, inverse Huffman coding, non-linear dequantization, inverse MDCT and subsequent Polyphase Filterbank, which mapped the subband frequencies back to the time domain in form of PCM samples of audio signal. Scale factors, which obtain form of side information, are used as multiplication factors for the dequantization step. Remaining part in form of ancillary data is trimmed in the process of decoding the input bitstream.

## IV. ALGORITHMS

### A. Encoding Watermarked Audio in Time Domain

This algorithm is based on embedding the watermark into a linear time domain (uncompressed audio) using a PCM watermarking and encoding afterwards. Both, the input signal and the output signal are uncompressed audio signals. Because encoding comes in following steps, algorithm of the inserting of watermark into uncompressed audio signal must be robust and inaudible enough into the audio. Algorithm proposed in advance is based on the SS modulation and satisfy these requirements. This algorithm is described below.

#### *Spread Spectrum*

Nowadays, from all the existing technologies dealing with the audio watermarking, the SS is the most popular technique for embedding watermarks into audio content. It features several advantages, for example high robustness against signal manipulations, high capacity and the attribute that the embedded watermark is distributed over the entire data.

Watermark is composed of the binary representation of the pseudorandom sequence (generated by a secret key) embedded into audio. The basic concept of SS watermarking is embedding a pseudorandom signal in the form of watermark with high bandwidth and low energy into original data. The output of this spreading operation is a watermarked signal with high bandwidth (usually 12–18kHz) [8].

In [7][8] authors present PCM watermark embedder based on the SS. After spreading a watermark, a perceptual model is applied to compute an estimate masking threshold. Masking threshold is used by a time-variant filter (adaptive or matched filter) in order to shape the energy distribution of the SS data according to the masking threshold. Resulting signal is added to the original audio signal, which can be in follows encoded by MPEG audio compression. After MPEG audio decoding step a watermark is detected using a matching filter with the filter coefficients matching the reserved spreading sequence, can be treated as a key. This approach has drawbacks because the MPEG algorithm is a lossy compression, therefore some of the watermarks bits are destroyed. This method is a fast scheme with low computational complexity. The time needed to PCM watermarking is sum with the time needed to MPEG encoding. It is useful in broadcasting and intellectual property protection.

### B. Partially Decoding-Watermarking-Reencoding

This algorithm is suitable primarily in the field of on-the-fly embedding. It operates the partially compressed domain. Watermarking is achieved by subsequent audio decoding, watermarking and reencoding. This generates decrease in sound quality audio signal and robustness of watermark.

Firstly, the bitstream is parsed by the bitstream multiplexer and divided into side information and then quantized and encoded by special values. Afterward the Huffman decoding and inverse quantization processes are applied to the spectral values in order to retrieve the spectral representation of the audio signal. Watermark is generated and spread in the bandwidth by using the SS modulation. After that the signal is converted into spectral representation of the encoder by means of the same analysis filter bank as used for audio

encoding alone. After spreading, the frequency spectrum of the watermark is shaped by applying an adaptive filter (coefficients are chosen dependent on the masking threshold). Resulting spectral data are added, quantized and coded in encoder part. The side information is used for these steps in order to avoid accumulation of quantization distortion. Output bitstream is generated by bitstream multiplexer [9]. Authors propose this method in the application of tracing illicit copies in the Internet by identifying the originator of each copy. Watermark can be added either into MDCT coefficients or in encoding step by Huffman entropy code. Disadvantages of this method are high computational complexity and loss of the quality of audio by multiple quantization. Following paragraph is the application of algorithm described above.

In the [10] authors proposed the technique so-called *ESC mode* based on the watermarking into *linbits* in the Huffman coding step in big value regions. Authors had described this method as follows. If the existing quantized coefficient values over 15 in this region, *ESC mode* starts to work. In *ESC mode*, the value of 15 is inserted into bitstream in the form of the value of the Huffman code table and the difference between the coefficient value and 15 is inserted into bitstream in the form of binary code as it is. The number of bits to code the differences is called the *linbits* and it can be maximum of 13 bits. Therefore although the bits are changed in the *linbits* part, MP3 file size does not change. In this case, authors evaluated good insertion rate for about 60 bytes per second and average MOS for about 4.6 [10].

### C. Watermark Insertion into Compression Encoder

The basic idea of this approach is to insert watermark directly during the encoding step. Uncompressed audio signal is processed by the analysis filter bank with parameters gathered from both, the signal itself and the perceptual model. Same parameters are used to transform the watermark that is obtained by spreading the watermark data using the SS modulation [8]. Watermark embedder does not have to use just the psychoacoustic model for computing masking threshold contained in MPEG encoder, it can also use an explicit masking model too. Weighting step by time-variant filter is applied to the watermark signal in order to obtain perceptual shaping. After shaping the watermark spectrum, both spectra are summed. Watermarked audio spectrum is quantized and coded afterwards. Information from the perceptual model is evaluated and frequency dependent scaling factors are calculated. This approach enables an optimal coordination between the quantization of the encoder and the watermark embedding process [11][12]. This algorithm has a good robustness and provides minimal degradation of audio quality and it is the fast algorithm, which combines encoding and watermarking in one step. Embedding a watermark is performed usually into MDCT coefficients.

### D. Watermarking in Compressed Bitstream

For embedding the watermark directly into compressed bitstream some methods based on watermarking into sample data or scale factors in the last years were proposed.

The algorithm designed in [4] embeds the watermark into the sample data. This method has drawbacks. Authors



describe fact, that if we change every sample by inserting a watermark, the distortion of resulting audio is easily detected by human ear, because the changing of these encoded samples is very sensitive. Authors randomly selected one or two samples to watermark, in order to solve this problem. By choosing the right samples, so-called the *spacing parameters*, the distortion can be minimized.

In the same article authors proposed another technique based on the insertion of a watermark by small change in the scale factors in every frame of compressed audio bitstream. Scale factors are multipliers that make the samples fully use the quantizer range. Each scale factor takes 6 bits therefore there exist as many as 63 levels of scale factors ( $2^6 - 1$ ). Authors indicate that the level changing of scale factors has an auditory effect. When scale factor level increase, the sound becomes stronger. In contrast, when scale factor level decrease, the sound becomes weaker [4]. Therefore, small change of scale factor level (1 or -1) normally can not be detected. Bigger changes produce perceivable audio distortions.

Both algorithms have more disadvantages. Embedding capacity and robustness are very small and they have a problem with multiple watermarks and resistance against signal manipulations is very small too.

#### E. Watermarking of Metadata

Metadata is data about another data. It can have the form of text data and in case of music material it contains the title, composer, player, genre, etc. as descriptive information about the actual content. In MP3 files, metadata is integrated in the ID3 tags (ID3v1 or ID3v2). For transmission of the watermark in practice are not usable because of the existence of enormous quantity of applications, which prove to edit the content of the ID3 tags and they are not part of the audio data. Another method is the insertion of the watermark into private bit in header of every frame in MPEG bitstream. This trivial method is impractical because the private bit is again not a direct part of the audio data. In this context, the Digital Rights Management (DRM) can also be mentioned. This technique for copyright protection is described in [13].

#### V. FUTURE WORK

The following research will be centred on testing selected methods, which insert the watermark in time domain, in the compression domain and during the compression procedure. The tests will be specialized on the robustness of algorithms against selected attacks and inaudibility of the inserted watermark using the subjective and objective criteria. The most suitable one of these algorithms will be then used in the proposed concept (see Fig. 3).

This system should be working on the on-line distribution of music files with inserted watermark. The system will generate a unique number (ID) only for a registered user. This number will be stored in customer database and used in the watermark's generation block. So generated watermark will be then inserted by the most suitable algorithm into files, which the users select by themselves from music database. Inserted music files will be as follows sent to the customer in compressed format. If such system was robust enough, it would have been used

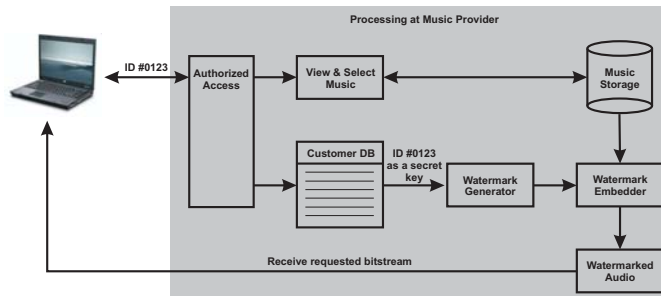


Fig. 3. The basic concept of pay audio system

also in commercial sphere for subscribed on-line distribution of music recordings protected by copyright, specifically by watermarks.

#### VI. CONCLUSIONS

Since human ear is more sensitive to audio distortion than human eye to image distortion, there exists a small amount of algorithms for embedding watermarks into compressed audio. Also the amount of data which can be embedded as watermark is very limited and depends on the content of the audio stream.

In the future we would like to target on designing the robust method with the use of spread spectrum watermarking for MP3 files, this is the methods for embedding watermarks into MDCT coefficients or Huffman coefficients. Therefore we would like to use the most suitable algorithm for designing the concept of the system for on-line music distribution.

#### ACKNOWLEDGMENTS

The work presented in this paper was supported by the Ministry of education of Slovak Republic under research projects AV 4/0006/07, VEGA 1/4054/07 and AV 4/2016/08 an Slovak Research and Development Agency under research project APVV-0369-07.

#### REFERENCES

- [1] D. Gruhl, A. Lu and W. Bender, "Echo hiding," *In Proceedings of the Workshop on Information Hiding*, Cambridge, vol. 1174 of Lecture Notes in Computer Science, pp. 295–315, May 1996.
- [2] J. Staš and J. Juhár, "A brief overview of digital audio watermarking algorithms," *Digital Technologies 2008*, Žilina, vol. 43, Nov. 2008.
- [3] W. Bender, D. Gruhl, N. Morimoto and A. Lu, "Techniques for data hiding," *IBM System Journal*, vol. 35, pp. 313–336, 1996.
- [4] L. Qiao and K. Nahrstedt, "Non-invertible watermarking methods for MPEG encoded audio," *Security and Watermarking of Multimedia Contents*, *In Proceedings of the SPIE 3675*, pp. 194–202, June 1998.
- [5] Davis Pan, "A tutorial on MPEG/audio compression," *IEEE Multimedia*, vol. 2, pp. 60–74, June 1995.
- [6] L. Gang, A.N. Akansen, M. Ramkumar and X. Xie, "On-line music Protection and MP3 compression," *IMVSP, Proceedings of 2001 International Symposium*, Hong Kong, pp. 13–16, 2001.
- [7] C. Neubauer and J. Herre, "Audio Watermarking of MPEG-2 AAC bit stream," *In 105th AES Convention*, Paris, Preprint 5101, Feb. 2000.
- [8] C. Neubauer and J. Herre, "Advanced watermarking and its applications," *In 109th AES Convention*, Los Angeles, Preprint 5176, Sep. 2000.
- [9] C. Neubauer, R. Kulesa and J. Herre, "A compatible family of bitstream watermarking schemes for MPEG-audio," *In 110th AES Convention*, Amsterdam, Convention Paper 5346, May 2001.
- [10] S.S. Yang, D.H. Kim and J.H. Chung, "Watermark insertion into MP3 bitstream using the *linbits* characteristics," *In 115th AES Convention*, New York, Convention Paper 5991, Oct. 2003.
- [11] F. Siebenhaar, C. Neubauer and J. Herre, "Combined compression/watermarking for audio signals," *In 110th AES Convention*, Amsterdam, Convention Paper 5344, May 2001.
- [12] F. Siebenhaar, C. Neubauer, J. Herre and R. Kulesa, "New results on combined audio compression/watermarking," *In 111th AES Convention*, New York, Convention Paper 5442, Sep. 2001.
- [13] S.H. Kwok, "Digital rights for the online music business," *ACM SIGecom Exchanges*, New York, vol. 3, pp. 17–24, 2002.

# Using Virtual Reality Environment for Modeling Software System

*Kristián ŠESTÁK*

Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

kristian.sestak@gmail.com

**Abstract**— This paper describes the importance of visual modeling in the stages of creating software and possibilities of using virtual reality environment for this purpose. Quickly completely and correctly understanding is crucial to every phase of the software development process. This process can be to incorporate into a virtual reality environment. Then, this model has benefits, which provide virtual reality environment, especially for user requirements specification, analysis and implementation into design models of the target system.

**Keywords**— methodology, software engineering, software life cycle, software modeling, software visualization, UML, virtual reality, virtual reality environment

## I. INTRODUCTION

The increasing complexity and robustness of software systems increases the demands placed on them, on understand. Creating a software system is a rather complex process, which consists of several stages. This process can be described as software life cycles. Each of these stages is important and has a significant influence on the resulting software product. This process represents the software life cycles. It is therefore important in this regard is still to improve the methodology of creating the software and adapt them to current trends.

The current trend in the field of the software modeling is to provide software visualization in three dimensions. Software visualization can be seen as a specialized subset of information visualization. Information visualization is the process of creating a graphical representation of abstract data.

We focus on research in the areas of Knowledge-based software life cycle and architectures [1].

## II. UNIFIED MODELING LANGUAGE

The Unified Modeling Language (UML) has established itself widely in the Software Engineering industry as the most used modeling language for Object-Oriented Analysis and Design (OOAD). UML is a 3rd generation object-oriented modeling language, unifies the best practice from Booch (Grady Booch), OMT (Jim Rumbaugh), OOSE (Ivar Jacobson), State charts (David Harel), Several dozen others. UML has been accepted by the Object Management Group (OMG) as a standard (1997) and UML is formally under development since 1994. [5, 6, 8]

### A. Deficiencies of UML

UML diagrams are composed from a small set of graphical primitives: basically, these consist of text, boxes, lines, and arrows. As a result, designers can easily draw UML diagrams by hand without colored pencils and stencils. [2]

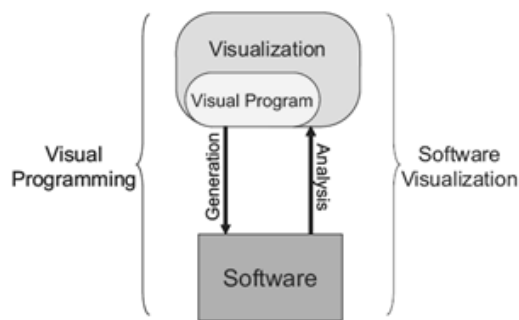
Despite the widespread use of this method, the visual efficiency of these diagrams is low. In recent user studies, so-called *geons* have been used to draw diagrams of software architectures. [Pourang et al. 2000]

The *geons* are a collection of 24 primitive, viewpoint-invariant 3D objects, which means that they are easy to recognize even when projected into 2D. Several experiments with computer science students showed that the subjects could visually analyze geon diagrams much faster and with more accuracy, and that they could recall them better in comparison with equivalent UML diagrams. [2]

In one experiment [Pourang et al. 2000], they first showed a UML or geon diagram for 15 seconds to the test persons, and then they presented them with diagrams of substructures and asked whether these occurred in the original diagram. All diagrams were shown on the computer screen and the test persons had just to press the “Y” or “N” key. In this experiment the average identification time was 4.3 seconds for geon diagrams, but 7.1 seconds for UML diagrams. But, even more importantly, the error rate was only 13% for geon diagrams, compared with 26% for UML diagrams. Thus the identification of substructures was much faster and more accurate with geon diagrams. In another experiment [Pourang et al. 2004], the test persons were given a problem description and four UML or geon diagrams. The task was to decide which one of the diagrams was created for that problem. In this experiment the average error rate was 15% for geon diagrams and 36% for UML diagrams. [2]

## III. SOFTWARE VISUALIZATION

Many authors define software visualization as the visualization of algorithms and programs. This definition excludes a lot of uses of visualization techniques in computer science and has also hindered synergies in the past. We define software visualization as the visualization of artifacts related to software and its development process. [2]



**Fig. 1 Visual programming versus software visualization**

As shown in Fig. 2, visual programming and software visualization complement each other. Software visualization generates visualizations from specifications of software systems, while visual programming generates software systems from visual specifications. [2]

The visualization can be concerned with visualizing: [2]

- Structure refers to the static parts and relations of the system, i.e. those, which can be computed or inferred without running the program. This includes the program code and data structures, the static call graph, and the organization of the program into modules.
- Behavior refers to the execution of the program with real and abstract data. The execution can be seen as a sequence of program states, where a program state contains both the current code and the data of the program. Depending on the programming language, the execution can be viewed on a higher level of abstraction as functions calling other functions, or objects communicating with other objects.
- Evolution refers to the development process of the software system and, in particular, emphasizes the fact that program code is changed over time to extend the functionality of the system or simply to remove bugs.

Software visualization is one approach suggested and being investigated worldwide for providing some assistance in program understanding. [14]

#### A. 2D versus 3D Representation

The work of Hubona, Shirah and Fout [Hubona et al. 1997] suggest that users' understanding of a 3D structure improves when they can manipulate the structure. Ware and Franck [Ware and Franck 1994] indicate that displaying data in three dimensions instead of two can make it easier for users to understand the data. In addition, the error rate in identifying routes in 3D graphs is much smaller than 2D [Ware et al. 1993]. The CyberNet system [Dos Santos et al. 2000] shows that mapping large amount of (dynamic) information to 3D representation is beneficial, regardless of the type of metaphors (real or virtual) used. Also, 3D representations have been shown to better support spatial memory tasks than 2D [Tavanti and Lind 2001]. In addition the use of 3D representations of software in new mediums, such as virtual reality environments, are starting to be explored [Knight and Munro 1999, Maletic et al. 2001]. [7]

The work of Gogolla [Gogolla et al. 1999] suggested several scenarios where UML diagrams can benefit from a

three-dimensional layout. They illustrated these scenarios by 3D UML diagrams that they implemented using the Virtual Reality Modeling Language (VRML): [2]

- In a class diagram, important classes are drawn in the foreground and thus have the focus. Moreover, one can have several different perspective views of the same diagram. In each of the perspectives the focus is on different classes, and when the user changes the perspective the nodes of the diagram are moved by smooth animation to their new positions.
- In object diagrams, various shapes can be used to represent objects. Objects of the same class have similar shapes.
- For sequence diagrams, animations show messages represented by small balls that move from the sender to the receiver. The problem of overlapping arrows for simultaneously sent messages and of arrows crossing lifelines can also be solved by 3D layout. Unfortunately, projecting these
- Several diagrams can be combined in space. For example, a class diagram may be shown in the background and a related sequence diagram in the foreground.

#### B. Visualizing Software in Virtual Reality Environments

The term “virtual reality” (VR) was first used in the 1980s. and has been defined by Hamit (1993) as “the presence of human in a computer generated space,” or more specifically, “a highly interactive, computer-based, multimedia environment in which the user becomes a participant with the computer in a virtually real world.”

Emphasizing the interaction and interface aspects, Stone regarded virtual reality in 1995 as an “interface between human and computerized applications based on real-time, 3D graphical worlds”.

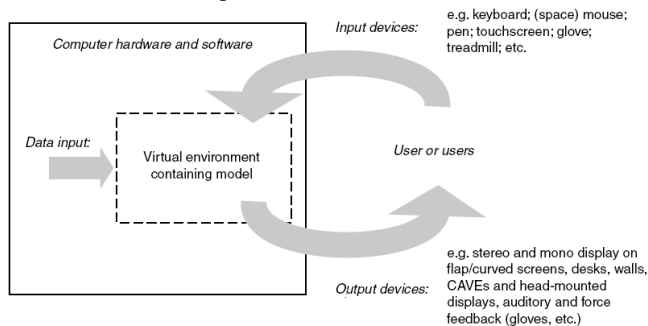
To extend the impact of realistic visualization and experience, Isdale defined virtual reality in 1995 as “a way for humans to visualize, manipulate and interact with computers and extremely complex data”. Such visualization is not limited to just graphics, but may also takes on a more general form of visual, auditory or other sensual outputs to the user.[3, 4]

According to these definitions, a virtual reality application has the following inherent important features. [9]

- Interactive: Realistic interactions with virtual objects via data gloves and similar devices to support the manipulation, operation, and control of objects in a virtual world.
- Real time: Viewing, interactions, and other related tasks have to be executed with real-time response so that the resulting illusion of being fully immersed in an artificial world is as convincing as possible.
- Immersive: Head-referenced viewing can be taken as an example to provide a natural interface for navigation in a 3D space, and can give the user the ability to look-around, walk-around, and fly-through in the virtual environment. Sound, haptic devices and other non-visual technologies can also be used to enhance the virtual experience significantly.

The components of a VR system are the computer hardware

and software, the input and output devices, the data and the users, as shown in Fig. 3.



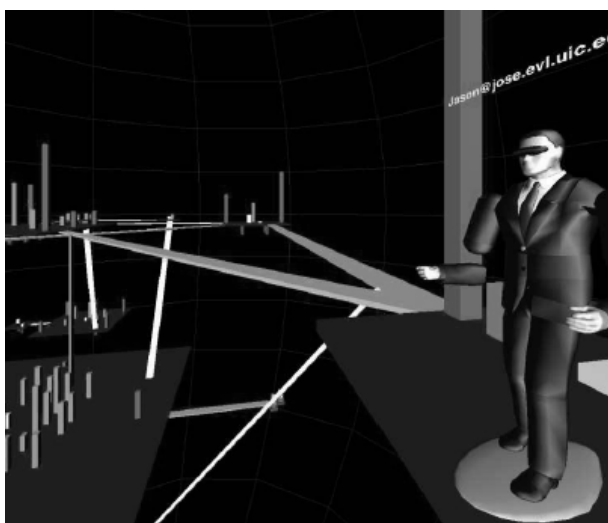
**Fig. 2** Components of a VR system

*C. Visualizing Software Systems*

*1) IMSOvision*

Authors [11] represent Visual language (built in VRML, OpenGL) Imsovision (IMmersive SOFTWARE VISualizatIOn) is a system that supports program understanding and development through software visualization. The building blocks of the language are:

- User centric (rather than compiler centric)
- Natural
- Support multiple abstraction levels
- Does not overload the user
- Maps naturally between abstraction levels



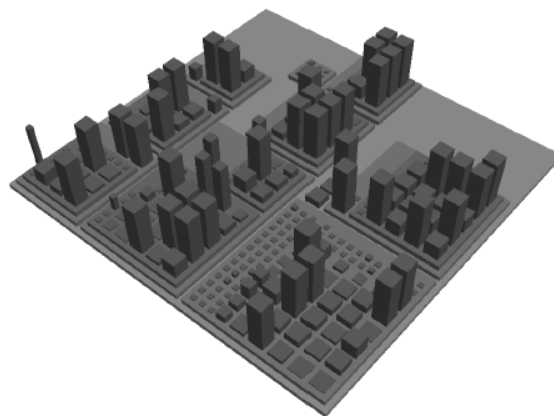
**Fig. 3** A remote user immersed in the VE investigating a visualization of a software system.

As shown in Fig. 5, Imsovision is a system that shows a stereoscopic 3D visualization of classes, their properties, and their dependencies and aggregations in a virtual-reality environment, namely a cave. As the cave is a room where the visualization is rear-projected onto its walls, the user can enter the room and interact with the 3D scene. In Imsovision, classes are represented by platforms, methods by columns, and attributes by spheres put on top of the platforms. The platforms of subclasses are placed next to their superclasses. Dependencies and aggregations are shown as flat edges. In addition, various properties are represented by colors. [11]

*2) Visualizing Software Systems as Cities*

Authors [15] represent classes as buildings located in city districts which in turn represent packages, because of the following reasons:

- A city, with its downtown area and its suburbs is a familiar notion with a clear concept of orientation.
- A city, especially a large one, is still an intrinsically complex construct and can only be incrementally explored, in the same way that the understanding of a complex system increases step by step. Using an all too simple visual metaphor (such as a large cube or sphere for the whole system or each package) does not do justice to the complexity of a software system, and leads to incorrect oversimplifications: We have to cope with the fact that software is complex.
- Classes and the packages they reside in are key elements of the object-oriented paradigm and thus, primary orientation point for developers. We currently do not display the class internals, because for a largescale understanding it is not necessary. Apart from over-plotting problems, it is also contrary to the way one explores a city: A person does not start the exploration of a real city by looking into particular houses. It is however necessary for a fine-grained understanding and thus a topic of our future research.

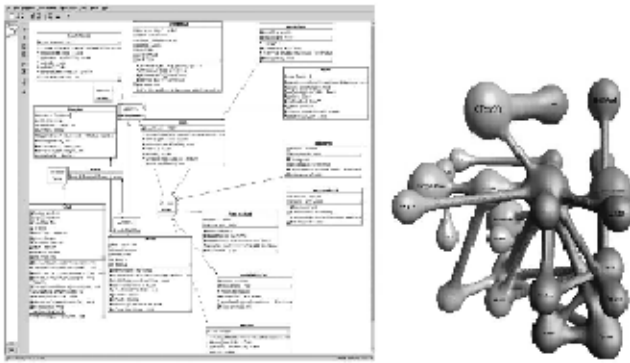


**Fig. 4** City of ArgoUML

Fig. 6 shows an isolated district. The classes are represented as buildings of the city and the packages as its districts. In [6] the authors also propose a city metaphor. In their case, the city represents a package and contains, for increased realism, non-source elements, such as trees, streets, and street lamps. In this metaphor, the program run would be represented as cars originated from different components specifically interested in developing systems which are effective on standard workstations. [15]

*3) Metaball metaphor*

Authors [14] represent particles in the metaball metaphor can be mapped to software structures, with spherical or elliptical blobs representing an object or a function (distinguished by different shapes for particles) that are created dynamically during a program execution and cylindrical blobs connecting these software entities for representing the interrelationship amongst them.



**Fig. 5 Combination of classical UML diagram with metaball visualization**

Fig. 7 shows a UML class diagram and a corresponding 3D metaball visual. For this example, the number of classes is small, and both visualizations appear to serve the intended purpose. But it is obvious that with limits on screen space the UML diagram does not scale. On the other hand, with the ability to navigate in 3D space and view the 3D structure from any viewpoint, the available space for pictorially depicting entities and relationships is vastly expanded. Further, combinations in which UML diagrams of metaball sub-clusters can be shown on demand would further enhance information communication. [14]

#### IV. CONCLUSION

Using 3D visualization of complex systems, we get better and easier to understand the visual model. Currently, UML gives us the 3D extension. Extensions in the form of plug-ins, but these are not directly integrated in the UML.

UML is not a methodology. UML has not been designed for a specific methodology. The methodology cannot be generalized there is no universal methodology which is suitable for all software systems. UML is the standard for modeling software systems. UML is independent of the software environment and extensible.

Visualization of the internal structure of software systems could be used for different purposes, but primarily, it is to support program comprehension.

Comprehension of object-oriented programs is simplified if the relationships that exist between classes and other parts of the program are easily understood. Diagrammatic notations that have evolved certainly help. Thus, UML based static and dynamic visualization techniques such as class models, sequence and collaboration diagrams can be applied for smaller software systems to provide an overview of the relationships in a program. However, for large software systems these diagrams do not provide adequate abstraction to visualize all the dependencies.

Creating intuitive and useful abstraction is one of the major research issues in 3D software visualization. Advantages of 3D for software visualization: 3D characteristics of entities can represent properties of underlying objects. The positions of objects in space may have relevance to properties. Program comprehension is an important part of not only software maintenance, but also the entire software engineering process.

One has to make distinction between 3D and VR. A user immersed in a Virtual Reality Environment (VE) can always access external information without leaving the environment

and the context of the representation. While both representations offer the perception of depth, only VEs allow the user to immerse oneself into the representation. Also, this immersion allows the user to take advantage of their stereoscopic vision. Stereopsis can be a great benefit in disambiguating complex abstract representations. It also helps the viewer to judge relative size of objects and distances between objects. In 3D, you have to move the view around to understand the diagram.

Given the complexity of software and the different problem solving characteristics of programmers, it is now well recognized that there is unlikely to be any one single visualization metaphor that can be considered most optimal for software visualization.

One of the main problems for software visualization is of trying to create a tangible representation of something that has no inherent form. Therefore the aim is to visualize the intangible in an effective and useful way. Effective and useful here refers to the visualization being able to increase the understanding of the user whilst reducing the perceived complexity. [13, 10, 6, 12, 11, 14]

#### V. ACKNOWLEDGMENT

This work was supported by VEGA Grant No. 1/0350/08 Knowledge-Based Software Life Cycle and Architectures.

#### REFERENCES

- [1] Havlice, Zdeněk et al. : Knowledge-based software life cycle and architectures. In: Computer Science and Technology Research Survey. Košice: TU, 2007. s. 47-68. ISBN 978-80-8086-071-4.
- [2] Diehl, S. Software Visualization: Visualizing the Structure, Behaviour, and Evolution of Software, Universität Trier Fachbereich Informatik 54286 Trier, Germany, 2007, ISBN 978-3-540-46504-1
- [3] Whyte, J. Virtual Reality and the Built Environment, 2002, ISBN: 0 7506 5372 8
- [4] Ko, Chi Chung & Cheng, Chang Dong: Interactive Web-Based Virtual Reality with Java 3D, National University of Singapore, Singapore, Copyright © 2009 by IGI Global, ISBN 978-1-59904-791-1
- [5] Turnbull, M.: An Overview of the Unified Modeling Language and Software Tools that Support it, CQUniversity Australia, Faculty of Sciences, Engineering & Health
- [6] Douglass, Bruce Powel: Real-Time UML, Chief Evangelist, I-Logix
- [7] Marcus, A., Feng, L., Maletic, J.I.: 3D Representations for Software Visualization, Kent State University, Department of Computer Science
- [8] Pender, T.: UML Bible, John Wiley & Sons © 2003, ISBN: 0764526049
- [9] Grady, Sean M.: VIRTUAL REALITY: Simulating and Enhancing the World with Computers, New Edition, Copyright © 2003, ISBN 0-8160-4686-7
- [10] Keown, L.: Virtual 3D Worlds for Enhanced Software Visualization, January 16, 2001
- [11] Maletic, J.I., Leigh, J., Marcus, A., and Dunlap, G.: Visualizing Object-Oriented Software in Virtual Reality, Division of Computer Science, Electronic Visualization Laboratory, The Department of Mathematical Sciences University of Illinois at Chicago, The University of Memphis
- [12] Knight, C.: System and software visualisation, University of Durham, Department of Computer Science, Visualisation Research Group
- [13] Knight, C. and Munro, M.: "Software Visualization conundrum", Department of Computer Science Technical Report 05/01, July 2001.
- [14] Rilling, J., Mudur, S.P.: 3D visualization techniques to support slicing-based program comprehension, Department of Computer Science and Software Engineering, Concordia University, 1455, de Maisonneuve West, Montreal, Que, Canada H3G 1M8
- [15] Wetzel, R. and Lanza, M.: Visualizing Software Systems as Cities, Faculty of Informatics - University of Lugano, Switzerland

# Robust video watermarking in DCT domain

<sup>1</sup>Tamás TOKÁR

<sup>1</sup>Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

<sup>1</sup>tamas.tokar@tuke.sk

**Abstract**—Digital watermarking has been proposed as a method for discouragement of illegal copying and distribution of copyrighted material. Currently proposed algorithms of digital watermarking in the context of the video are robust against unintentional attacks, but many of them do not protect enough against malicious attacks. This paper presents a new frequency domain-based video watermarking technique with improved robustness. Finally, achieved results are also presented, which verify the enhanced robustness of proposed system against malicious and non-malicious attacks.

**Keywords**—Digital watermarking, discrete cosine transform, frequency domain, video, watermark.

## I. INTRODUCTION

Multimedia production and distribution nowadays is digital. The World Wide Web, digital networks and multimedia offer unlimited opportunities to abuse copyrighted material. Digital watermarking is an appropriate tool for multimedia content protection. This technique consists in hiding of additional information within multimedia data, which is represented by a watermark. Digital watermarking has first been extensively studied for still images. Today however, many new watermarking schemes are proposed for other types of digital multimedia data such as audio and video. The most popular application of digital watermarking is copyright protection, where the purpose of the embedded watermark is to prove intellectual ownership. In the context of video there are other applications of digital watermarking such as copy control, video authentication, broadcast monitoring, fingerprinting, enhanced video coding etc. [1].

The watermark embedded into video must meet some requirements depending on target application whereby the most important are robustness, transparency and data payload [2]. Robustness means resistance of watermarking schemes against malicious and non-malicious attacks. It can be also seen as the ability of the detector to extract the hidden watermark from altered labeled video. The robustness is often evaluated via the survival of the watermark after attacks. Transparency is requirement for perceptual invisibility of the additional information. Watermark embedding brings some distortions into video and it causes quality degradation that should remain imperceptible for a human observer. The data payload is the amount of side information (watermark), i.e. the number of bits that can be embedded into video data without significant degradation of original quality.

In this paper a robust watermarking scheme is presented that embeds watermark invisibly and resists against chosen intentional and unintentional attacks.

## II. SYSTEM DESIGN

The introduced watermarking scheme is based on watermark embedding in frequency domain using discrete cosine transform (DCT). Discrete cosine transform is widely used in signal processing. Most of compression standards for video encoding such as MPEG perform 2D DCT on 8×8 pixel blocks of video frames. The resulting DCT coefficients in 8×8 blocks are consequently quantized with 8×8 quantization table (matrix) and entropy coded [3]. The main idea of the embedding algorithm is the required relationship enforcement between chosen DCT coefficients. The watermark embedding is performed frame-by-frame whereby the same watermark information is embedded into each video frame. Some intentional attacks on watermark in video such as frame dropping, frame swapping or averaging can cause temporal desynchronization of the watermark detector. Due to this desynchronization many watermarking schemes fail during watermark detection procedure, mainly when they require original video for watermark extraction. The proposed scheme does not require the knowledge of the original video during watermark recovery and therefore the threat of the detector's temporal desynchronization is eliminated.

### A. Watermark embedding

The embedding algorithm is performed in frequency domain by modification of certain DCT coefficients. The whole watermark embedding process is shown in Fig. 1.

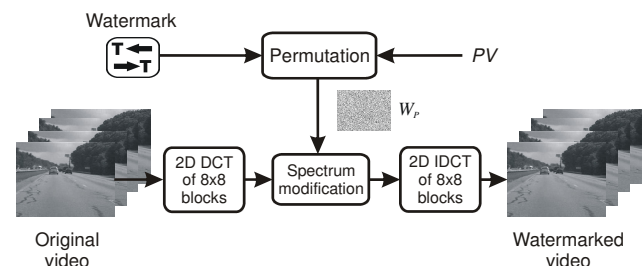


Fig. 1. Watermark embedding process

First the video frame in spatial domain is divided into 8×8 non-overlapping blocks. For each block a 2D DCT resulting 8×8 blocks of DCT coefficients, is then applied. The original watermark is pre-processed by permutation that randomizes watermark bits by using a permutation vector  $PV$ . The purpose of permutation operation is to give noisy nature for the watermark, which makes it undetectable. Further purpose of permutation is security enhancement of the proposed

scheme. Permutation vector  $PV$ , as part of the secret key keeps information about original positions of watermark bits. The noise-like permuted watermark  $W_p$  is then embedded into video frame.

Some algorithms in image watermarking exploit relationships between DCT coefficients within the  $8 \times 8$  block for embedding one watermark bit. However these relationships can be changed due to quantization because different DCT coefficients within  $8 \times 8$  blocks are not quantized the same way. On the other hand compression standards for image and video encoding use the same quantization matrix for all  $8 \times 8$  blocks of frequency coefficients independently of quantization step [4]. This means that coefficients at the same location in all  $8 \times 8$  blocks are quantized identically and therefore the relationship between them is leaved after quantization, too [5].

For this reason the proposed scheme embeds each watermark bit into 4 adjacent  $8 \times 8$  DCT blocks as shown in Fig. 2. by modifying one of certain coefficient in the first block  $B_1^{(s)}$ . Superscript ( $s$ ) refers to embedding of  $s$ -th bit of the permuted watermark  $W_p$ .

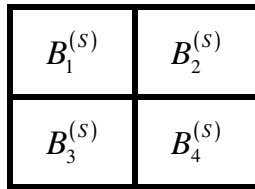


Fig. 2. Four adjacent  $8 \times 8$  blocks of DCT coefficients

The modification of chosen DCT coefficient in order to hide one watermark bit can be formally described as follows:

$$\begin{aligned} \text{if } W_p(s) = 1 \text{ and } K_1(x, y) < \bar{M} + \alpha \text{ then} \\ \Rightarrow K_1(x, y) = \bar{M} + \alpha \\ \text{if } W_p(s) = 0 \text{ and } K_1(x, y) > \bar{M} - \alpha \text{ then } \\ \Rightarrow K_1(x, y) = \bar{M} - \alpha \end{aligned} \quad (1)$$

where  $K_1(x, y)$  means the chosen mid-band coefficient to be modified with coordinates  $x, y$  in the first DCT block  $B_1^{(s)}$ . Number  $\bar{M}$  represents the mean value of coefficients with the same position in another three DCT blocks. Factor  $\alpha$  indicates the modification intensity and its value depends on the required watermark robustness. Mean value  $\bar{M}$  is defined in the following form:

$$\bar{M} = [K_2(x, y) + K_3(x, y) + K_4(x, y)]/3 \quad (2)$$

where  $K_2, K_3, K_4$  are frequency coefficients with the same  $x, y$  coordinates as  $K_1(x, y)$  in DCT blocks  $B_2^{(s)}, B_3^{(s)}, B_4^{(s)}$  respectively. Values of coordinates  $x = 3, y = 3$  were experimentally determined with respect to achievement of good perceptual transparency of the watermark. With such a modification (1) the desired relationship between  $K_1$  and  $\bar{M}$  is enforced.

After embedding of all watermark bits the  $8 \times 8$  block IDCT (*Inverse discrete cosine transform*) is applied on the modified frequency spectrum in order to obtain watermarked frame in spatial domain. Note that the same embedding algorithm is applied on all video frames separately, whereby each frame is labeled with the same hidden information, i.e. the watermark is embedded redundantly into video.

### B. Watermark extraction

The watermark extraction process is performed in frequency domain, too. Due to the presence of the same watermark information in all video frames, the watermark can be extracted from arbitrary frame without the knowledge of the original unmarked video. The extraction process starts with  $8 \times 8$  block DCT of watermarked frame (Fig. 3).

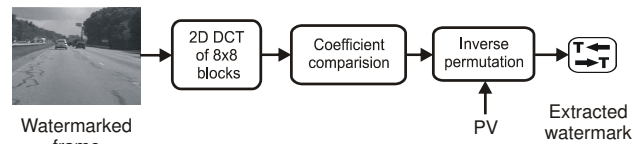


Fig. 3. Watermark extraction process

Information bit recovery is based on relationship between frequency coefficients at the same position of 4 adjacent DCT blocks. The hidden bit is simply extracted by comparison of chosen coefficient  $K_1$  in the first DCT block with mean value  $\bar{M}$  of coefficients in 3 another DCT blocks. The extraction of  $s$ -th watermark bit can be described as follows:

$$\tilde{W}_p(s) = \begin{cases} 1 & \text{if } K_1(x, y) > \bar{M} \\ 0 & \text{if } K_1(x, y) < \bar{M} \end{cases} \quad (3)$$

where  $\tilde{W}_p$  represents the extracted noise-like watermark, which is subsequently undergone to inverse permutation. The same permutation vector  $PV$  used at embedding serves to reconstruct the final watermark  $\tilde{W}$  from extracted information  $\tilde{W}_p$ . Note that extraction process is performed faster than embedding process because of sufficiency of one watermarked frame only for watermark recovery.

## III. EXPERIMENTAL RESULTS

The goal of the realized experiments was to verify the imperceptibility of the watermark and its robustness against intentional and unintentional attacks. The visual quality of the watermarked frames was evaluated, too. Frame dropping, frame averaging and frame swapping were tested as intentional attacks [6].

Simple frame dropping means removing one or several watermarked frames from the video sequence, frame swapping causes order modification of 2 or more consecutive frames. At frame averaging attack the mean of 2 or 3 consecutive frames is computed and these frames are replaced with the resulted averaged frame.

Video encoding standards of different types (MPEG-1, MPEG-2, MPEG-4/AVC, MJPEG) that introduce lossy compression of video data were used as unintentional attacks. Higher bit rate during video encoding reduces data volume at the cost of video quality degradation. According to ISO/IEC

recommendation the bit rate range is from 1 Mbps up to 2 Mbps for MPEG-1 video (typically 1,4 Mbps) and from 4 Mbps up to 20 Mbps for MPEG-2 video. The MPEG-4 / AVC compresses video at half bit rate with equivalent quality as in MPEG-1 or MPEG-2. Motion JPEG (MJPEG) belongs to first video encoding algorithms and it applies JPEG compression for each video frame individually. The compression level in standard MJPEG is usually given by quality factor  $Q_f$  (typical value  $Q_f = 85\%$ ).

In the experiments video with frame resolution  $320 \times 240$  and a binary logo as watermark with size  $20 \times 15$  were used. Note that video with grayscale frames is represented in true color mode i.e. all color planes (R-red, G-green, B-blue) of the frame are identical.

Table I shows experimental results after various attack types for different values of factor  $\alpha$ . Peak signal-to-noise ratio (PSNR) refers objective measure for evaluating of the visual quality of the watermarked frames. An example of original video frame and corresponding watermarked one is shown in Fig. 4. Bit equality (BE) in percents represents the amount of watermark bits correctly recovered from the watermarked video frame.

TABLE I

TEST RESULTS FOR VARIOUS INTENTIONAL AND UNINTENTIONAL ATTACKS

Intensity factor		$\alpha = 5$	$\alpha = 15$
PSNR [dB]		49,47	45,81
Attack kind	Attack type	Bit Equality [%]	
Unintentional	MPEG-1 at 1.4 Mbps	90,13	100
	MPEG-2 at 4 Mbps	87,29	100
	MPEG-4 / AVC at 700 kbps	79,68	93,51
	MJPEG ( $Q_f = 85\%$ )	68,52	99,56
Intentional	Frame dropping	100	100
	Frame swapping	100	100
	Frame averaging	100	100
Combined	Compression + Frame averaging	94,13	98,57

The results in Table I indicate high robustness of the proposed approach against chosen attacks whereby the embedded watermark has good perceptual transparency. Note that higher intensity factor increases robustness, but distortions in watermarked frames causing visual quality degradations are more perceptible and PSNR has lower values. The watermark is perfectly extracted at intentional attacks due to different video frames contain the same hidden information.



(a) (b)

Fig. 4. Subjective quality of original (a) and corresponding labeled frame (b)

#### IV. CONCLUSIONS

In this paper a DCT domain-based video-watermarking scheme was proposed. As experimental results show the proposed algorithm is highly robust against various types of attacks. The main advantage of the system is that watermark extraction process does not require knowledge of the original unmarked video. Due to this fact the system is also robust against malicious attacks that cause temporal desynchronization of the watermark detector.

The disadvantages of the proposed approach are relatively high computational complexity due to DCT and the need of uncompressed (or decompressed) video for watermark embedding. For these reasons the proposed system is not suitable for applications (e.g. broadcast monitoring) where watermark embedding or detection must be performed in real time. However some other applications such as copyright protection, fingerprinting or copy control do not require real-time watermarking. The proposed scheme is appropriate for labelling of stored videos such as DVD films or TV movies that are not encoded and broadcasted in real time. The watermark detection and extraction processes are performed in relatively short time because their algorithm does not process whole watermarked video, but only one or several watermarked frames. In future work some modifications of the proposed algorithm are needed, which reduce its computational costs.

#### ACKNOWLEDGMENT

The work presented in this paper was supported by Grant of Ministry of Education and Academy of Science of Slovak republic VEGA under Grant No. 1/4054/07.

#### REFERENCES

- [1] G. Doerr, J.-L. Dugelay, "A guide tour of video watermarking," In *Signal Processing: Image Communication 18*, Sophia-Antipolis (France), 2003, p. 263–282.
- [2] M. Wu, B. Liu, *Multimedia Data Hiding*. Springer-verlag, New York, Inc. (USA), 2003.
- [3] D. Levický, *Multimediálne telekomunikácie*. 1. vyd. Košice ELFA s.r.o., 2002, 240s, ISBN 80-89066-58-5.
- [4] ISO/IEC 13818-2, Generic Coding of Moving Pictures and Associated Audio, Recommendation H.262 (MPEG-2), 1995, International Standard
- [5] C. Y. Lin, S. F. Chang, "A robust image authentication method distinguishing JPEG compression from malicious manipulation," In *IEEE Transactions on circuits and systems for video technology*, vol. 11, no. 2, 2001, pp. 153-168.
- [6] P. -W. Chan, *Digital Video Watermarking Techniques for Secure Multimedia creation and Delivery*. PhD thesis, The Chinese University of Hong Kong, 2004.



# Contextual adaptation in sound localization: temporal aspects

Beáta TOMORIOVÁ, Rudolf ANDOGA, Michal BARTO, Norbert KOPČO

Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

beata.tomoriova@tuke.sk, rudolf.andoga@tuke.sk, michal.barto@gmail.com, norbert.kopco@tuke.sk

**Abstract**—Experiment was performed to measure the effect of auditory context with various temporal characteristics on horizontal sound localization. Subjects' task was to localize a target sound presented from either left or right hemifield of a frontal horizontal plane. Context was represented by the fact that target sound was sometimes preceded by a distractor sound presented from the position directly ahead of the listener. Temporal parameters of the context such as frequency of occurrence of the distractor or distractor-to-target onset asynchrony were parametrically varied during experiment. Results showed that perceived positions of target sounds were shifted away from distractor position due to context. Shifts were slightly larger for higher frequency of occurrence of distractor and for the largest and shortest distractor-to-target onset asynchrony.

**Keywords**—contextual adaptation, auditory plasticity, sound localization.

## I. INTRODUCTION

Mechanisms of human sound localization are based on differences in time, level and spectrum between how is the same sound heard in each ear [1]. Specific values of these differences (also called localization cues) by activating certain neurones in auditory processing pathway give rise to perception that sound comes from a specific position in space. However, the mapping between localization cues and perceived position is not static. It changes with anatomical changes of the head (during development) [2] or can be artificially induced by occluding one ear [3] or by other mechanical means. These examples are related to relatively long period of time such as days or weeks.

Except for this long-term plasticity, a short-term plasticity operating on time scales of minutes was observed in some studies. It was shown that localization of a sound can be influenced by preceding “distracting” sound with longer duration. For example, if a constant sound lasting approximately 4 minutes - so-called “adaptor” was presented to subjects, they perceived subsequent target sounds as shifted from its actual position in direction away from the adaptor [4].

Similar effect was also observed in [5]. In this study the trials that contained just sounds to be localized (target sounds) were interleaved with trials that contained both distractor and target sound. Even though the distractor was in this study identical to target sounds and had the same duration, shifts in localization were still observed. Authors suggest that

“contextual adaptation” could have been induced, where context was represented by presence of trials with distractor.

## II. EXPERIMENT

### A. Motivation and hypotheses

Results of study [5] outlined that localization could be affected by context since shifts in localization were observed. However, the size of the effect could not be estimated as the experimental procedure was originally designed to study just the effect of directly preceding distractor on subsequent sound and shifts in perceived positions of targets which weren't directly preceded by distractor were unexpected.

Current study based on previously mentioned is designed to examine the effect of the context. Specifically, it focuses on how the adaptation depends on temporal aspects of the context such as:

- frequency of occurrence of distractor
- stimulus onset asynchrony (SOA) between distractor and target (time between the onset of the distractor and onset of the target sound).

It also examines temporal profile (build-up and decay) and spatial profile (dependence on the position of target sound) of the contextual shifts.

Main hypotheses are:

- contextual shifts will be stronger for higher frequency of occurrence of distractor
- shifts will have similar magnitude for all target sound positions (based on results from previous study [5])

### B. Methods

Experiment was conducted in a sound-proof booth of 3 x 2 x 3.1 meters. Subject was seated at the center of a quarter

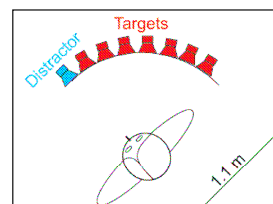


Fig. 1. Experimental setup. Nine loudspeakers were equally spaced along the quarter circle. Just 8 loudspeakers are used in each block of trials (depicted). Loudspeaker ahead of the listener was aimed for presentation of distractor, other loudspeakers except the most-sided (which is used as a distractor when subject is oriented to other side) presented target sounds.

circle of equally spaced 9 loudspeakers (Fig.1). The radius of the circle was 1.1 m.

Subjects' task in the experiment was to localize a target sound by pointing in perceived direction of the target with a hand-held pointer. The responses were scanned by a videosystem. Subjects were instructed to have their eyes closed during the experiment to prevent visual feedback.

Target sound stimulus was a 2-ms frozen noise burst presented randomly from one of the seven middle loudspeakers (Fig. 1). The target was sometimes preceded by identical distractor sound presented from loudspeaker directly ahead of the listener (the left-most or right-most from Fig.1 depending on subject's orientation). Trials where target was preceded by distractor are referred to as "distractor trials" and represent the context aimed to induce plasticity. Trials with target alone are referred to as "no-distractor trials" and enable to deduce how localization was affected by context.

Experiment consisted of four sessions. Each session contained 8 types of blocks, each of 140 adaptation trials + 14 preadaptation and 35 postadaptation no-distractor trials, which enable to study build-up and decay of adaptation.

In adaptation part of each block an adaptation was induced by specific type of context which was fixed within a block and defined by one combination of:

- percentage of no-distractor trials within a block: 50%, 25% or 10% (indirectly indicating also the frequency of the occurrence of distractor)
- SOA between distractor and target: 25 ms, 100 ms or 400 ms

After each block subjects changed orientation (facing left-most resp. right-most loudspeaker) to prevent possible inter-block effects when testing the same hemisphere.

One of the blocks in each session contained just no-distractor trials. Since no plasticity was experimentally induced here and therefore localization in this block was supposed to be "normal", it represented the baseline for analysis of the effect of the context.

The effect of the context could be therefore estimated by comparing performance in "baseline block" with performance in no-distractor trials from distractor blocks. Comparison was made by estimating biases, which were for each type of context computed by subtracting mean perceived azimuth in a block with specific type of context from mean perceived azimuth in baseline block.

### III. RESULTS

Biases up to 4 degrees relative to actual position of target were observed for each type of context (Fig. 2 and Fig. 3). The direction of the biases was away from distractor. The differences between effects of different types of context are small, however, biases tend to slightly increase for context with lower percentage of no-distractor trials within a block (bars with same color but different intensity). Biases are also larger for shortest and longest SOAs (blue and red bars relative to green).

Results also showed that the biases are larger for those target positions, which are close to distractor position (Fig. 3).

Adaptation built up relatively quickly, approximately within first 2 minutes after the first onset of the distractor (Fig. 4).

During the rest of the adaptation part biases remained relatively constant and after the offset of the distractor they decreased.

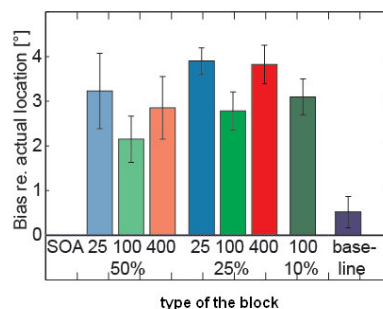


Fig. 2. Bias in responses relative to actual target location, averaged across target locations. Across-subject mean and within-subject standard error.

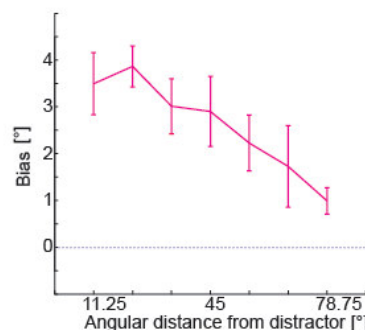


Fig. 3. Bias in responses relative to baseline block, averaged across target location and context type. Across-subject mean and within-subject standard error.

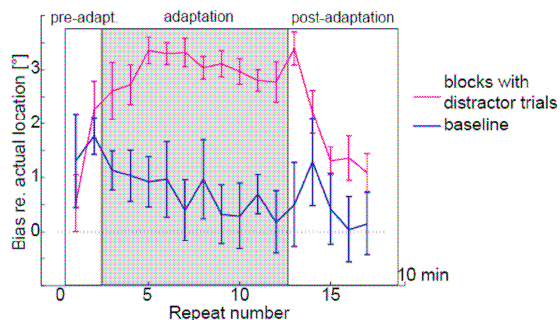


Fig. 4. Build-up and decay of adaptation shifts as a function of repeat number within a run (number of times the target was presented from specific position), averaged across context type and target location. Across-subject mean and within-subject standard error.

### IV. DISCUSSION

Results of the current study indicate that sound localization can be affected by context, which induces shifts in perceived position of target sound.

The shifts had magnitude up to 4 degrees, direction away from distractor and were slightly dependent on the frequency of occurring of the distractor. The higher was the frequency (resp. percentage of distractor trials) the larger shifts were

observed. One possible explanation is that if distractor caused some changes in neural representations of auditory space lasting few seconds, then higher frequency of its occurrence would affect more subsequent trials. Another explanation, from a top-down point of view, is that higher frequency of the distractor could change subject's strategy – since the probability that some trial would contain distractor is higher, subjects could during the experiment actively focus their attention away from distractor position to make the localization of target easier. This strategy could be efficient especially in case of context with SOA = 25 ms, where distractor and target are very proximate in time and localizing a target can require some effort.

Consistent with previous explanation, biases for shortest SOA are large. However, biases caused by the longest SOA have almost the same magnitude (while middle SOA lower). The reason for this effect is not clear.

Spatial profile of adaptation indicates that biases for targets presented near distractor position are larger than for other targets. Reason could be that neural representations of spatial positions, which are near the distractor position, are affected more. On the other hand, since the responses in baseline block are also shifted, there is a possibility that another way of responding (not using pointer) would be more appropriate and the effect wouldn't be so strong.

The build-up and decay of the adaptation was relatively quick. It is not clear whether the adaptation was caused by the top-down processes (changes due to strategy, active focusing) or bottom-up processes (changes in neural representations due to the presented stimuli).

Another experiments need to be performed to describe the effect of the context on localization. Except for temporal properties of the context, spatial properties such as dependence on position of the distractor could be studied. Other experiments could focus on effect of way responding, or estimation whether the contextual adaptation is bottom-up or top-down, etc.

The research in spatial auditory perception and especially auditory plasticity can be useful in field of prosthetics. Cochlear implants can partially restore sense of hearing to deaf people by electrically stimulating the auditory nerve. However, the auditory space perception has been distorted in deaf patients and studying the ways how to restore it after the implantation by training (or how to remap auditory space in general) could lead to improvement of their auditory perception.

#### ACKNOWLEDGMENT

This work was supported by NIH #1R03TW007640 and VEGA #1/3134/06.

#### REFERENCES

- [1] YOST, W.A., (2000), *Fundamentals of hearing: An introduction* (4th ed.), New York: Academic Press.
- [2] MOORE, D.R. & KING, A.J., Plasticity of binaural systems. In T.N. Parks, E.W. Rubel, R.R. Fay & A.N. Popper (eds.) *Plasticity of the Auditory System*, Springer, New York (2004)
- [3] KING AJ, PARSONS CH, MOORE DR: Plasticity in the neural coding of auditory space in the mammalian brain. *Proc Natl Acad Sci, USA* 2000

- [4] CARLILE, S., S. HYAMS, et al. (2001). Systematic distortions of auditory space perception following prolonged exposure to broadband noise. *J Acoust Soc Am* 110: 416-425.
- [5] KOPCO, N., BEST, V., SHINN-CUNNINGHAM, BG (2007) Sound localization with a preceding distractor, *J. Acoust. Soc. Am.* 121, 420.

# MUDOF Meta-Learning Algorithms for Automatic Selection of Algorithms for Text Classification

<sup>1</sup>Gabriel TUTOKY

<sup>1</sup>Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

<sup>1</sup>gabriel.tutoky@tuke.sk

**Abstract**—This paper presents a meta-learning approach for textual document classification task and an automatic selection of the best available algorithm for creation of classifiers. After brief introductory description of principles of creation and evaluation of the classifiers, the meta-learning approach is presented as a method for automatic selection of the most appropriate classifier algorithm for creation of binary classifiers. Designed methods, based on the modification of MUDOF (Meta-learning Using Document Feature Characteristics) algorithm, are described together with its implementation using the JBoWl (Java Bag of word library). Finally, the experimental results achieved by the meta-learning algorithms as well as their comparisons with traditional ways used for text classification are presented.

**Keywords**—Error! Reference source not found..

## I. INTRODUCTION

The text classification, also sometimes referenced as the text categorization, is a method for data analysis from texts [1]. It is based on the supervised learning, where the goal is to distribute the textual documents from input data collection to the pre-defined categories. The input data collection contains a sub-set of training examples, i.e. the documents categorized in advance; these training examples are processed by statistical or machine-learning algorithms to produce the so-called classification model. The resulting model can then be applied on the rest of the input data collection to classify the textual documents without known relation to the categories [2].

Classification of text documents was originally designed as a semi-automatic procedure, where the users (usually experts) were responsible for selection of proper classification model, algorithms, text pre-processing methods, and optionally also to restrict the training set. In the most of application, process of semi-automatic text classification is unusable, because users are not experts in the field of text mining. It is hard for them to select the optimal settings and the requirement was to try to investigate the classification settings automatically, from global characteristics of the input data collection. It resulted in a design of the meta-learning method for automatic selection of classification algorithms. This method can be used in applications where is needed the classification of textual documents and where is impossible to require any optimal setting for classification process from users.

## II. TEXT CLASSIFICATION, BASIC PRINCIPLES

The classification belongs to one of the basic approaches in predictive data mining. In the case of text classification, it is an approach for specific knowledge extraction from textual documents. The process of classification consists of two phases [1]:

1. Construction of the classifier
2. Usage of the classifier

Basic functional blocks and components used in these two phases are depicted on Figure 1.

In the *first phase*, a given set of training examples (i.e. a set of already categorized text documents) is processed to create the classifier as a model of the data behavior. In the *pre-processing* step, the terms are extracted from the text of documents, and the whole input set is transformed into a vector representation [2]. The vector size can be reduced by various *pre-processing* and text analysis methods as e.g. tokenization, stop-words elimination, stemming and lemmatization, term clustering (LSI), etc. [2], [3].

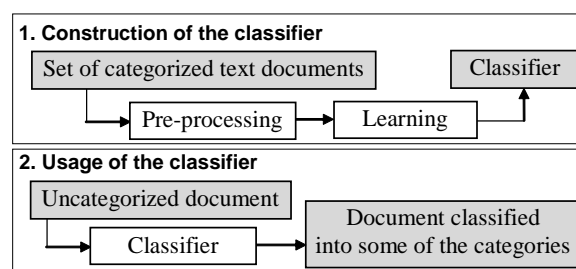


Fig. 1. Two phases of the classification process

In the step of *learning*, various learning algorithms based on the statistical and heuristic techniques are used for processing the vector representation of the training set. Selection of proper algorithm and settings its optimal parameters is usually preformed manually and requires expert knowledge as well as an experience in the field of text mining. This is especially the point where the meta-learning approach (described in the next section) can help and select the most appropriate classification algorithm with respect to the characteristics of the training data set. In this paper are focused following algorithms:

- Linear classifiers: *Perceptron*, *Support Vector Machine (SVM)*,

- Methods based on a recursive division of the space of documents into a set of disjunctive areas: *Decision trees*, *Decision rules*,

- Methods based on the instances: *k-Nearest Neighbors* (*kNN*).

For the construction of the classifier, it is assumed that every training example belongs to one or more of the pre-defined categories  $c_i$  ( $c_i \in C$ ,  $C = \{c_1, c_2, \dots, c_N\}$ , where  $N$  stands for the number of the categories). Because the document can be categorized into more than one category, the problem is decomposed into individual category level. Specifically, there is one classifier, so-called *binary classifier*, for each category. Binary classifiers are able to distinguish the documents of one category from the documents belonging to all of the rest of categories.

This way, each category can have its own binary classifier; decompose the classifier building problem and allows using different types of classifiers for various categories. The union of these binary classifiers for all categories forms the resulting classifier – so-called *classification model*, which implicitly describes the set of pre-defined categories.

The resulting classification model is used in the *second phase* for classification of "new" (i.e. unknown, a-priori uncategorized) documents from the input data collection. The input document is processed by all the binary classifiers from the classification model and document is assigned into these categories, for which the binary classifier has a positive value, e.g. if binary classifier  $CF_a$  classify input document as positive, then the document is assigned into category  $c_a$ . If there is no binary classifier for input document which returns the positive value, then the document is assigned as unclassified. Finally, set of categories, classified for the input document, is generated as a result of the classification procedure.

*Quality* of the classification can be evaluated using the *testing data collection* of documents, which contains the documents already (a-priori) categorized into the pre-defined categories. The testing documents are classified regularly, using the produced classification model (Fig. 1). The results are then compared with the a-priori categorization for each testing document. This comparison is performed by a set of statistical measures; the most frequently used indicators are the *precision*, *recall*, and combined *effectiveness measure F1*. These measures can be combined into one global measure for the space of all categories by *micro averaging* and *macro averaging* methods [2], [3]. These measures will use to evaluate the results of experiment in section IV.

### III. META-LEARNING

Implementation of the classification procedure in practice requires the selection of proper algorithm in the learning step. The meta-learning approach can be used to automate the selection of the algorithms separately for each of binary classifiers, according to the specific characteristics of the training set of documents. This approach does not require any additional effort from user side for controlling the classification process and provides higher quality of the classification results.

The meta-learning approach is based on a design of an adaptive system, which can increase its effectiveness based on

the feedback from previous "experiences", i.e. on the evaluation of the examples processed in past [4]. Selection of the best learning strategy, most suitable for particular problem, is a generalization based on accumulating experience on the performance of multiple applications, strategies, or algorithms [5]. In the domain of text classification, the meta-learning approach is able to select the most appropriate and the most effective classification algorithm according to the characteristics of the training set (as e.g. term or category distribution, average length of documents, etc.). To achieve this selection, there is a need to create the decision mechanism (meta-model) in the first step and then to use it in the second step for creation of new classifiers (cf. first phase of general text classification process, presented on Fig. 1).

The process of the meta-learning approach applied in text classification for construction of classifiers consists again of the two phases, as depicted on Fig. 2:

1. *Construction of the meta-model*;
2. *Usage of the meta-model* for selection of algorithms and for creation of classifiers.

*First phase* of the meta-model construction can be further divided into the two steps:

- Specification of feature characteristics for training documents;
- Learning of the meta-model (meta-classifier).

The feature characteristics can be obtained from the training set for each category and can be expressed for category  $c_i$  as a vector  $F_i = (f_{i1}, f_{i2}, \dots, f_{in})$ . The vectors of all categories can then be used in the step of meta-model learning for modeling and future selecting the most appropriate algorithm for particular categories.

The meta-model learning is originally based on prediction of an optimization parameter [6], given by comparison and evaluation of the feature characteristics  $F_i$  with the values of effectiveness, i.e. with the classification errors obtained from applying pre-defined classification algorithms on the training and testing set. The meta-model can then be constructed from these values using a regression analysis.

*Second phase* of the meta-model usage is rather simple, where the feature characteristics are obtained from unknown (uncategorized) input documents and these are processed in the same way as in the phase of meta-model construction. The meta-model is then able, according to the feature characteristics of the new documents, to select and propose the most suitable classification algorithm for creation of the resulting classifier.

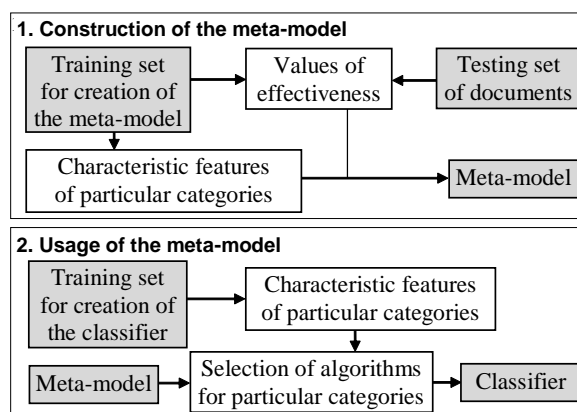


Fig. 2. Meta-learning approach, two phases

In my work, I adopted the MUDOF algorithm [6], based on the multiple regression analysis of feature characteristics obtained from the training set of text documents. I have implemented the MUDOF algorithm as an extension of the JBow library [7]. In addition, after a set of initial experiments, I have enhanced the MUDOF algorithm itself in two ways. The first way consist of small modification of original MUDOF algorithm (signed as MUDOF\_R and implemented into to JBow) [8] and second way is producing of a new, modified version, of MUDOF algorithm, where the meta-model learning step is based on the meta-classification procedure, using the kNN classification algorithm.

The modified algorithm, referenced as MUDOF\_K (i.e. MUDOF with kNN meta-learning) was also implemented into the JBow. A set of experiments were performed to compare the effectiveness and results of the MUDOF\_R (i.e. MUDOF based on regression analysis) with MUDOF\_K and with the traditional classification using the five pre-defined algorithms (see section 2). Some of these experiments are described and discussed in the section 4 below.

The original MUDOF algorithm proposes a set of nine future characteristics [6], from which we have selected five and modified it into the form of a ratio or average value [8]:

- *AvgTopInfoGain*, average information gain of the best  $t$  terms of a given category. The information gain of individual terms is computed for current category, average is then counted from  $t$  terms with the highest information gain.

- *PosTr*, ratio of positive and negative examples in the training set for given category.

- *AvgTermVal*, average weight of document's terms for given category. The average weight of terms for a single document is computed at first; then the weight is computed for all the positive examples of a given category.

- *NumInfoGainThres*, ratio of the number of terms with the information gain over the threshold to the number of all the terms.

- *AvgDocLen*, average length of a document for given category. The document's length is computed as a number of all the indexed terms in a document. The average is obtained by computing the length for all the positive examples for given category.

Both MUDOF algorithms require a division of the training set into two sub-sets [6]:

- training set for meta-model ( $TM$ ),
- training set for classification model ( $TC$ ).

The feature characteristics of the particular categories are obtained from  $TM$  and  $TC$  as two separate data sets. The values of feature characteristics obtained from  $TM$  are used in the *first phase* for creation of the meta-model. MUDOF\_K algorithm forms the meta-model so that the feature characteristics of particular categories are represented as points in the  $|F|$ -dimensional feature space in the meta-model. After creation of the meta-model are memorized  $|C|$  points of which each one represents of one category in the space. The values of memorized points are  $|A|$ -dimensional vectors of values of effectiveness, where  $A$  is the set of base algorithms.

In the *second phase*, the feature characteristics obtained from  $TC$  are used for selection of the most proper algorithms for creation of the final classifier. The selection is performed by searching through feature space in the meta-model

individually for each category (for each vector of feature characteristics) from  $TM$ . For one category, the  $k$ -nearest neighbors (points in future space of the meta-model) are founded on the basis of similarity function<sup>1</sup>. For each of  $k$ -founded points is known which algorithm is the best for memorized point (category). The algorithm with the highest occurrence from  $k$ -founded points is selected for creation of binary classifier. If the count of highest occurrences is equal for two or more algorithms, the algorithm from the most similar point is selected (also on the basis of the same similarity function as above).

Our implementation of the MUDOF\_K algorithm can be described in the following steps:

---

#### A. Meta-model construction:

Input:  $TM, TC$ , set of available classification algorithms  $A$ , set of categories  $C$ .

1. For each (category  $c_i$  from  $C$ )
2. Compute  $F_i$  for  $c_i$  from  $TM$
3. While (there is an algorithm in  $A$ )
4. Take an algorithm  $ALG_j$  from  $A$
5. Apply  $ALG_j$  on  $TM$  for  $c_i$  and obtain the binary classifier  $CF_{ij}$
6. Apply  $CF_{ij}$  on  $TC$  for  $c_i$  and obtain the optimization parameter  $p_{ij}$
7. End While
8. Save  $F_i$  and corresponding vector of optimization parameters  $P_i = (p_{ij}, \dots, p_{i|A|})$
9. End For

#### B. Usage of the meta-model:

10. For each (category  $c_i$  from  $C$ )
  11. Compute  $F_i$  for  $c_i$  from  $TC$
  12. Compute distances  $d_{ik}$  between  $F_i$  and all  $k$  points in the meta-model using *similarity function*
  13. Find the  $k$ -nearest points toward  $F_i$
  14. Obtain occurrences of particular algorithms for founded points
  15. ALG with the highest occurrence is the best for category  $c_i$
  16. End For
- 

Main advantage of the proposed modification is the possibility of incremental learning of the meta-model. This feature is especially helpful in the systems, where the input data set is updated rather frequently and the changes should be reflected in the meta-model.

## IV. EXPERIMENT

The meta-learning algorithm MUDOF\_R, based on the regression model, as well as the MUDOF\_K modification, based on the  $kNN$  classification method, were both implemented as an extension of the *JBowl* library. The implementations were then tested to prove the concept of automatic creation of classifiers by the meta-learning approach and to evaluate the quality of the resulting classification procedure.

#### A. Preparation of the testing data

The experiments were accomplished on the *Reuters-21578* [9] set of documents. It contains 10.788 documents distributed into 90 categories. For the experiments, the set of documents was divided into the following subsets:

- training set ( $TR$ ): 7.769 documents,
- testing set ( $TE$ ): 3.019 documents.

For the meta-learning, the  $TR$  was further divided into the training sets for meta-model and for classifier:

<sup>1</sup> In our case, the similarity function is used Euclidean distance function

- *TM*: 3.815 documents,

- *TC*: 3.961 documents.

The Reuters-21578 set is not very well balanced; it has a high variability of the documents distribution towards the categories. It contains categories with about 1.500 positive examples, as well as about 30 categories with less than 10 documents.

### B. Executing of experiment

The experiment was focused on testing of the meta-learning approach on a Reuters-21578 data set. The goal was to prove the hypothesis that the meta-learning provides an automatic selection of algorithms with the best possible (or better) effectiveness and quality of the resulting classifier in comparison with the several pre-defined classification algorithms.

The *effectiveness* of the classification was evaluated by the *F1 quality measure* mentioned in the section 2 above. The integrated measure *Macro F1*, which combines precision and recall over whole testing set, was used as the main quality measure for the experimental results. The MUDOF\_K and MUDOF\_R algorithms were compared with basic classification algorithms as *Decision Trees*, *Decision Rules*, *SVM*, *Perceptron*, and *kNN*. Resulting values of the quality measures are listed in Table I., graphical comparison of the *Macro* measures is depicted on Fig. 3. The *Macro* measures have been chosen because they are the most descriptive effectiveness measures for unbalanced document.

TABLE I. QUALITY OF MEASURES

Statistics	MUD-OF_K	MUD-OF_R	Dec. Trees	Dec. Rules	SVM	Perc.	kNN
<i>Micro Precision</i>	0,808	0,869	0,790	0,792	<b>0,932</b>	0,885	0,852
<i>Micro Recall</i>	<b>0,860</b>	0,820	0,793	0,801	0,785	0,794	0,792
<i>Micro F1</i>	0,833	0,844	0,792	0,796	<b>0,852</b>	0,837	0,821
<i>Macro Precision</i>	0,567	0,556	0,521	0,499	<b>0,580</b>	0,556	0,496
<i>Macro Recall</i>	<b>0,520</b>	0,502	0,503	0,492	0,369	0,356	0,384
<b><i>Macro F1</i></b>	<b>0,543</b>	0,527	0,511	0,495	0,451	0,434	0,433

The results demonstrate that the MUDOF algorithms, using the meta-learning approach, are able to provide automatic selection of algorithms additionally with higher values of the resulting effectiveness, expressed by the Macro F1 measure. For the macro measure, the MUDOF has similar results as Decision Trees and Rules. However, the MUDOF has better results for the Micro measure. The SVM, Perceptron, and kNN have similar and slightly better (in case of SVM) results as MUDOF for the Micro measures, but the MUDOF is better in the results for Macro measures. Percentage increase of Macro F1 measure of the MUDOF algorithms in comparison with the globally best basic algorithm, i.e. Decision Trees, was 3,2% for MUDOF\_K and 1,6% for MUDOF\_R algorithm.

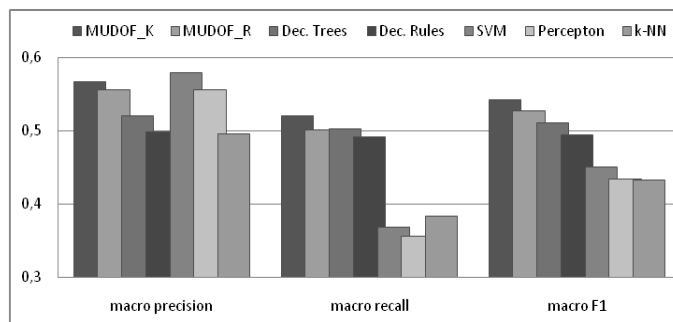


Fig. 3. Comparison of the Macro measures for algorithms

## V. CONCLUSION

The presented meta-learning approach towards the text classification seems to be a suitable method for support of automatic classification in user-oriented systems. The original MUDOF meta-learning algorithm, based on the linear regression, was modified and adapted using the kNN classification method for meta-model creation. Both algorithms were tested on the Reuters-21578 set of documents and the results indicate that the meta-learning provides an automatic selection of algorithms and increases effectiveness and quality of the results. However there is still some space for further improvements of meta-learning algorithms. After all, the proposed meta-learning approach can be considered as a technology, which enables automatic and adaptive text classification, increases quality of the classification results, and can be effectively used in the user-oriented systems in practice.

## REFERENCES

- [1] Paralič, J.: Knowledge Discovery in databases and texts, Habilitation thesis, Technical University of Kosice, Slovakia, 2003.
- [2] Bednár P.: Automatic classification of texts based on the content (in Slovak), Concept of PhD. thesis, TU Košice, Slovakia, 2004.
- [3] Sebastiani F.: Machine Learning in Automated Text Categorization, ACM Computing Surveys, Vol. 34, Iss. 1, New York, USA, 2002, pp. 1-47.
- [4] Vilalta R., Drissi Y.: A Perspective View And Survey Of Meta-learning, AI Review, Vol. 14, No. 2, Springer Netherlands, 2002, pp. 77-95.
- [5] Vilalta R., Giraud-Carrier Ch., Brazdil P.: Meta-Learning: Concepts and Techniques, The Data Mining and Knowledge Discovery Handbook, Springer US, 2005, pp. 731-748.
- [6] Wai L., Kwok-Yin L.: A meta-learning approach for text categorization, Proc. of the 24th ACM SIGIR conference, New Orleans, USA, 2001, pp. 303-309.
- [7] Bednár P.: JBowL, Java Bag of word library, available at <http://sourceforge.net/projects/jbowl/>, Accessed: 12<sup>th</sup> May 2008.
- [8] K. Furdík, J. Paralič, G. Tutoky: Meta-learning for Automatic Selection of Algorithms for Text Classification, Proc. of the conference 19<sup>th</sup> Central European Conference on Information and Intelligent Systems, Varaždin, Croatia, 2008, pp. 477-484
- [9] Lewis D.: Test data Collection Reuters-21578, available at <http://www.Daviddlewis.com/resources/testcollections/reuters21578/>, Accessed: 12<sup>th</sup> May 2008.

# Linear logic proof search as a stochastic game

<sup>1</sup>Anita VERBOVÁ

<sup>1</sup>Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

<sup>1</sup>anita.verbova@tuke.sk

**Abstract**—In this contribution we present stochastic linear logic proof game. The winning strategies of the Player are the proofs of a given formula. The multiplicative and additive connectives of linear logic are described by means of probabilistic operators. We represent proofs by the sequences of game triangles. We introduce first the deterministic version of the linear logic proof game. Afterwards we present a stochastic version of this game and illustrate how stochastic game can be used for the proof search in linear logic.

**Keywords**—linear logic, proof search, stochastic game, probability.

## I. INTRODUCTION

Linear logic proof search may be seen as a game. This game is called *the linear logic proof game* and it is played on linear logic formulae, and its moves are instances of inference rules of linear logic. This kind of game enables us to work with stochastic games formulated by linear logic. Connections between linear logic and probabilistic games considered in complexity theory are investigated in [1].

In this paper we present also the stochastic linear logic proof game, because we are interested in introducing probability to linear logic. We achieve this by the linear logic proof game, where Nature tosses a coin and he can choose with the probability  $\frac{1}{2}$  some branch of the proof. It depends exactly on this choice how the game continues. Consequently we do not need to pass through all possible branches of the proof tree, therefore we need less time for the verification of the provability of a sequent.

During the proof search, there are non-deterministic choices which come from the additive connectors. Accordingly, there is some probability of choosing one way in the proof tree, which we want to describe by our stochastic linear logic games.

## II. LINEAR LOGIC

Linear logic [2] provides a logical perspective on computational issues such as the control of resources and the order of evaluation. Before we show an example of a game, we explain some basic notions necessary for understanding the connection between linear logic and game theory. MALL is the multiplicative-additive fragment of linear logic, which allows only multiplicative and additive connectives and their units with negation. Logical connectives of MALL are: multiplicative connectives  $\otimes$  (conjunction),  $\wp$  (disjunction) with units  $1, \perp$ , resp. and additive connectives  $\&$  (conjunction),  $\oplus$  (disjunction) with units  $\top, 0$ , resp. We described logical connectives and their inference rules in [3], [4]. In the presented games we use these rules in the form of linear sequent calculus. Here we present only these rules, which we will use in this paper. Let in the following  $p$  be a proposition (literal),

$A, B$  be linear logic formulae and  $\Gamma, \Delta$  be finite sequences of linear logic formulae.

$$\frac{}{\vdash p, p^\perp} \text{ (identity)} \quad \frac{\vdash \Gamma, A \quad \vdash A^\perp, \Delta}{\vdash \Gamma, \Delta} \text{ (cut)}$$

$$\frac{\vdash \Gamma, A \quad \vdash B, \Delta}{\vdash \Gamma, A \otimes B, \Delta} (\otimes) \quad \frac{\vdash \Gamma, A, B, \Delta}{\vdash \Gamma, A \wp B, \Delta} (\wp)$$

$$\frac{\vdash \Gamma}{\vdash \perp, \Gamma} (\perp) \quad \frac{}{\vdash \top} (\top)$$

$$\frac{\vdash A, \Gamma}{\vdash (A \oplus B), \Gamma} (\oplus 1) \quad \frac{\vdash B, \Gamma}{\vdash (A \oplus B), \Gamma} (\oplus 2)$$

$$\frac{\vdash A, \Gamma \quad \vdash B, \Gamma}{\vdash (A \& B), \Gamma} (\&) \quad \frac{}{\vdash \top, \Gamma} (\top)$$

## III. GAMES

Games are used to model interaction between a System and its Environment. One of the players in the game is taken to represent the System, and is referred to as Player; the other represents the Environment and is referred to as Opponent. A single play represents the interaction between the Player and the Opponent. It will be represented by a sequence of moves.

*Strategies* are rules specifying the way how the Player should play [5]. Formally we define the *winning strategy* for a player to be a function that chooses the moves in a way that he can win definitely.

A *symmetric game* is a game where the Player wins with a particular strategy dependently only on the strategies employed by the Opponent, not on who is playing them. That means that the change of players does not influence the winning possibilities. Most commonly a game is symmetric if the players have the same set of strategies to choose from, and the utility does not depend on the position of the players.

In *asymmetric games*, the two players fill up different roles (e.g. there is a stronger and a weaker player). The roles are known by both players (for example, they both know which of them is stronger). The roles may affect the possibility of winning or even the available strategies, but such differences may also be absent. Asymmetric games are games where there are not identical strategy sets for both players.

Asymmetric games are used to characterize different computational problems. We will use these games for the description of linear logic proofs. We follow the approach introduced in [6], [7], [8] with some modifications, desirable for our approach.



IV. LINEAR LOGIC PROOF GAME

We introduce linear logic proof game, which represents a proof search in a fragment of linear logic (MALL). This game belongs to asymmetric games, because in our game each player has different roles and these are known by both of them.

For the purposes of linear logic proof game we want to represent MALL inference rules as moves in this game. We are reasoning about a plane figure, which has to satisfy the following claims:

- 1) It must be able to represent one conclusion and one or two assumptions of each inference rule.
- 2) It must have such a shape that it could be connected with each other, so that one assumption of the previous figure is adjacent to the conclusion of the next figure.

We found it suitable to choose triangles for inference rules, because they have 3 edges, which could represent one conclusion and at most two assumptions. Therefore we partitioned our game triangles in a way, which is illustrated on Fig. 1. These game triangles are steps of proofs of a linear logic sequent or as moves in the linear logic proof game. Every triangle can be regarded as an inference rule, where conclusion of the rule is in the bottom part and assumptions (one or two) are in the left and right parts. The game triangle for axiom has in the bottom part only literals and information about winning or loosing the game ("WIN"/"LOSE").

There exist two players: the Prover and the Verifier. Both have given their own proof triangles. Formally the Prover's goal is to play a sequence of its game triangles proving a MALL sequent. The role of the Verifier is to force the direction of the Prover's proof in a way that makes it impossible for the Prover to win.

The Prover plays a sequence of triangles. The last two game triangles for the Prover from Fig. 1 have special role for continuing the game. He wins, when the game ends with scoring "WIN".

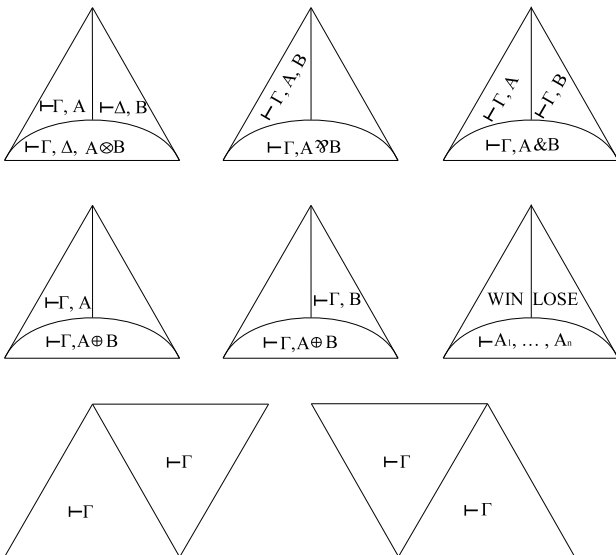


Fig. 1. Game triangles for the Prover

The role of the Verifier is to force the way of proof built by the Prover into the direction, where the Prover can not win. He plays with dark triangles illustrated in Figure 2. His major task is to block one side of the Prover's triangle. The Verifier has the knowledge whether a multiset of literals is an axiom,

so he can score the game by blocking "WIN" or "LOSE" in the Prover's triangle, leaving the correct score exposed.

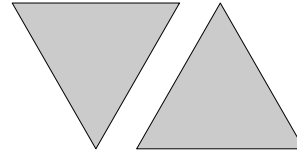


Fig. 2. Blocking game triangles for the Verifier

If an inference rule has two assumptions, Verifier can block one of them and to force Prover to continue in the unblocked way.

The literal triangle "WIN/LOSE" is used to score the game. The Prover plays until there are no connectives remaining in the sequent. After that the Verifier compares the literals in the last triangles. If the literals represent an axioms, the Verifier may block only "LOSE" edge. In this case the Prover wins. Otherwise, the Verifier blocks WIN edge, i.e. the Prover loses and the Verifier wins.

In short we can say that the Prover wins if and only if the sequence of triangles represents the legal proof of a given formula.

The Prover plays from the bottom of the initial triangle and builds a proof tree. He plans the layout of the proof so that all triangles are placed by the rules. Sometimes he must use connector triangles, which do not change the sequent within the current branch. They only extend a branch to reach the part of a board where is more space.

The Verifier checks whether all triangles are placed according to defined rules. In other words, he verifies whether multisets of formulae at adjoining edges are identical. Furthermore he checks whether all uncovered edges are edges of a literal triangle and whether this unblocked upper triangle should be scored as winning. That means, he checks all leaves of a proof tree. He allows the Prover to win if and only if every leaf is an axiom.

This form of a game may require more game triangles than the number of connectives in a formula, because the inference rule for & has more symbols in the assumptions and duplicates all of the context of a formula.

We present the fundamental principle on a simple example (Fig. 3), which illustrates a proof of the formula  $\vdash b, a \oplus (b^\perp \otimes a^\perp), a$ , where  $a, b$  are propositions.

This game is as if it was one-sided, because it seems that the Prover plays the game alone. The Verifier does not have any influence on the play of the game. He only checks the moves and scores all the branches of the game. In this kind of a game the whole proof tree is set on the board. This game is not very interesting for us, but we extend it and create a new game from it.

V. ALTERNATING LINEAR LOGIC PROOF GAME

We modify deterministic form of the linear logic proof game can be played according to the following rules, so that we get a new game alternating linear logic proof game.

The Prover tries to show that a sequent is provable. He can win if a sequent is provable. He cannot use connector triangles in this version of game.

The Verifier is able to prevent the Prover from winning if sequent is not provable. Both players are able to understand

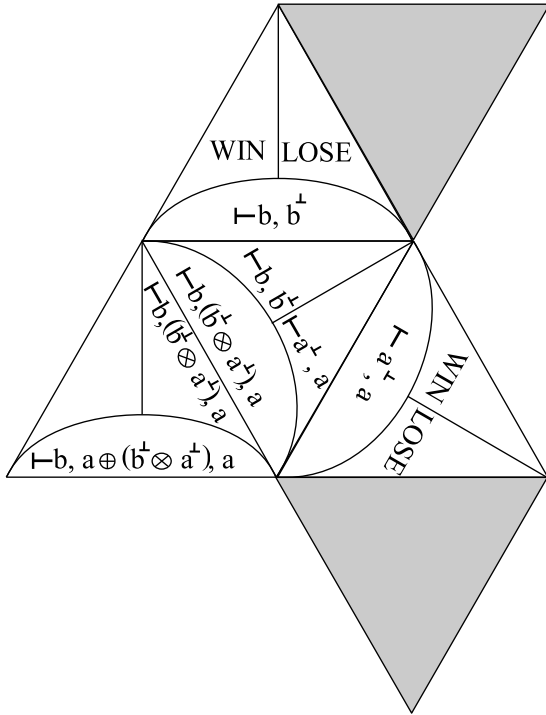


Fig. 3. Deterministic game, where plays only the Prover and the Verifier only scores the game.

provability of MALL formulae. A difference in comparison to previous games is that the players *alternate*.

The meaning of game triangles in this kind of a game is explained as follows. In that case if a *game triangle* has two sequents in the upper part, then the Verifier must block one of them. Therefore the Prover must continue by putting his triangle to the remaining unblocked sequent. If there is only one sequent in the upper parts of the game triangle then the Verifier blocks the empty edge and then the Prover is on turn.

It is possible here, that the Prover wins over an unprovable sequent (illustrated on Fig. 4) as a result of the Verifier's suboptimal but legal choice in the case of the  $\otimes$  triangle. In that case one subsequent is chosen by the Verifier, which is already provable ( $\vdash a^\perp, a$ ). If the Verifier blocks this side  $\vdash a^\perp, a$ , then he will win and the Prover will lose.

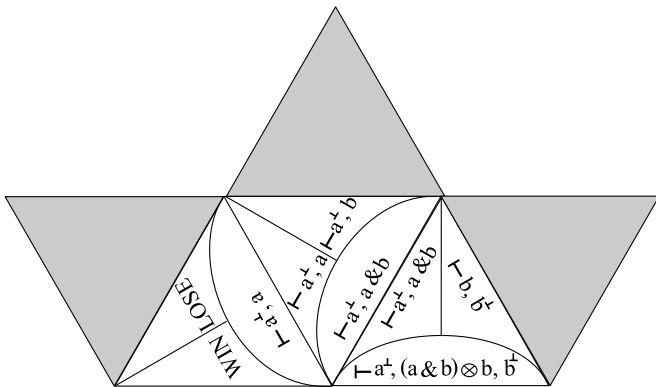


Fig. 4. Deterministic game, where players alternate. Only one branch of the proof is represented on the board.

We construct *only one* branch of a proof tree, because the Verifier must always choose only one way for the Prover to continue. If the Prover begins with a provable sequent, then he wins all the time, because all branches can be completed. If a

formula is an unprovable sequent then Verifier can force the Prover to the position, where the Prover has no further moves. Then the game is over and Verifier wins, because there remains a triangle which is not an axiom and is scored by Verifier as "LOSE". The length of a game is bounded by the number of connectives in initial sequent, while every move of the Prover reduces it by one and the Verifier always blocks one edge of the triangle.

A disadvantage of this kind of a game is that both players must have equal computational capacities.

If the Prover starts with the sequent  $A \& B$ , where  $A$  and  $B$  are composite formulae and one of them is provable, but the other not, then Verifier can win, if he is able to decide which is provable.

It follows from the nature of the linear connective  $\&$ , where one of the formulae  $A$  and  $B$  is performed, but we do not know which. As a real game, it is not effective, because both players can decide already from the first triangle placed on the game board, whether they have a winning strategy or not.

This game requires exactly the same number of game triangles as the number of connectives in the formula, because the Verifier always blocks one side of the triangle or puts his blocking triangle next to an empty edge.

## VI. STOCHASTIC LINEAR LOGIC PROOF GAME

In game theory, stochastic game is a competitive game with probabilistic transitions played by two players. At each step, the next move is made by both on the basis of the strategies followed by the individual players as well as the element of chance.

In this *stochastic linear logic game* the Prover is stronger and the Verifier is weaker player. This means, that the Verifier can only toss coins to choose a branch of a proof tree in the case of the inference rule for the connective  $\&$ .

This form of game has the feature that the Prover with knowledge can win with probability more than  $\frac{1}{2}$  if he starts with a provable formula and win with probability less than  $\frac{1}{2}$  with an unprovable formula.

A difference compared to the alternating linear logic proof game is that the Verifier does not block one sequent in the upper triangles in the case of  $\otimes$ . The Prover must complete more than one branch of a proof tree. He can use connector triangles to prevent from blocking on the board.

On Fig. 5 we illustrate *game triangles* representing all possible moves in our stochastic linear logic proof game.

If the Prover begins with an unprovable sequent, then there exists at least one branch of a proof tree which does not lead to an axiom. The number of branches depends on the number of connectives  $\&$ . In this case the Verifier chooses one of these branches with probability  $\frac{1}{2}$ . The Prover has a good chance of winning, because the probability that Verifier chooses a branch, which does not lead to an axiom, is very small. The chance of winning for the Prover and the Verifier in the case of  $\&$  we demonstrate on a simple example.

The Prover begins with a sequent  $A, A^\perp \& B$ . The Verifier can choose between  $A, A^\perp$  and  $A, B$ . If he chooses randomly, than the Prover has a 50% chance of winning. This means, the Verifier scores literals  $A, A^\perp$ , which represent an axiom if  $A$  is a proposition. The Prover has also 50% chance of losing the game if Verifier has to score literals  $A, B$ , which are not an axioms. The Verifier chooses "WIN" or "LOSE"

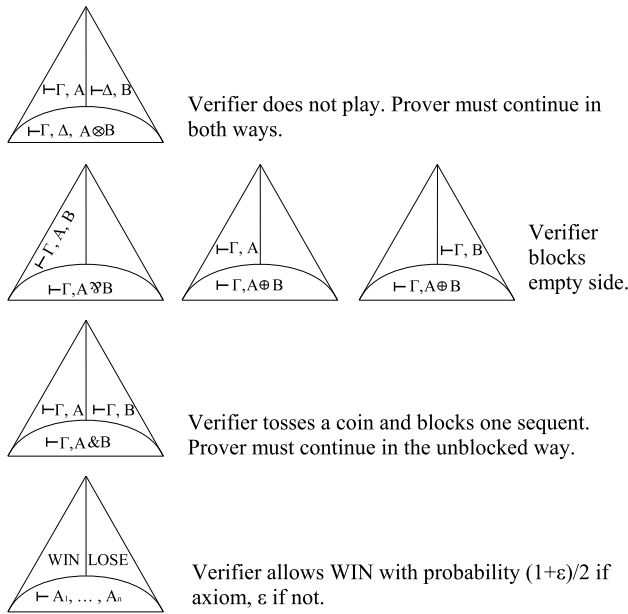


Fig. 5. All possible moves of both players in stochastic linear logic game

deterministically according to literals. The Prover has the same chance of winning or losing over an unprovable sequent.

If we want to give the Prover a good chance of winning ( $> \frac{1}{2}$ ) only for provable sequents and we do not want the Verifier to have any insight whether the initial sequent is an axiom or not, then we should introduce probability in the case of literal triangle. If the branch does not end with an axiom, it is more certain that the sequent is not provable. Therefore we can assign a lower probability of winning.

### VII. CONCLUSION

In our paper we consider linear logic as a logic, where formulae are actions. Formulae in linear logic can be regarded as resources that are exhausted after their use. Therefore this logic is appropriate for describing processes similarly as in real life.

The main objective of this contribution is the definition of *stochastic linear logic proof game*. We were interested in the implementation of probability in linear logic. By introducing probability in the game we define stochastic game of linear logic. We can use this game semantics in linear logic proof search, because winning strategies in the game represent proofs of a sequent.

### ACKNOWLEDGMENT

This work was supported by VEGA Grant No.1/0175/08: Behavioral categorical models for complex program systems.

### REFERENCES

[1] P. Lincoln, J. C. Mitchell, and A. Scedrov, "Linear logic proof games and optimization." *Bulletin of Symbolic Logic*, vol. 2, no. 3, pp. 322–338, 1996. [Online]. Available: <http://dblp.uni-trier.de/db/journals/bsl/bsl2.html#LincolnMS96>

[2] J.-Y. Girard, "Linear logic." *Theoretical Computer Science*, vol. 50, pp. 1–102, 1987.

[3] A. Verbová, V. Novitzká, and D. Mihályi, "Unification of proofs and verification of the correctness of proof structures," in *Analele Universitatii din Oradea, Proceedings of the 9th International Conference EMES*, vol. 9. Polish Information Processing Society, 2007, pp. 132–137.

[4] A. Verbová, V. Novitzká, and V. Slodičák, "From linear sequent calculus to proof nets," in *Informatics 2007, Proceedings of the Ninth International Conference on Informatics*. Slovak Society for Applied Cybernetics and Informatics Bratislava, 2007, pp. 100–107.

[5] S. Abramsky, "Semantics of interaction: an introduction to game semantics," in *Proceedings of the 1996 CLiCS Summer School, Isaac Newton Institute*, P. Dybjer and A. Pitts, Eds. Cambridge University Press, 1997, pp. 1–31.

[6] S. Abramsky and R. Jagadeesan, "Games and full completeness for multiplicative linear logic," in *Foundations of Software Technology and Theoretical Computer Science (FST-TCS'92)*, 1992, pp. 291–301. [Online]. Available: [citeseer.ist.psu.edu/article/abramsky94games.html](http://citeseer.ist.psu.edu/article/abramsky94games.html)

[7] A. Blass, "A game semantics for linear logic," *Annals Pure Appl. Logic*, vol. 56, pp. 183–220, 1992, Special Volume dedicated to the memory of John Myhill.

[8] P. Lincoln, J. C. Mitchell, and A. Scedrov, "Optimization complexity of linear logic proof games," *Theor. Comput. Sci.*, vol. 227, no. 1-2, pp. 299–331, 1999.

# Visualization and evaluation of ontological models (March 2009)

<sup>1</sup>Jozef VRANA

<sup>1</sup>Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

<sup>1</sup>jozef.vrana@tuke.sk

**Abstract**—Ontologies have been proven useful in wide range of areas. In the past a massive number of ontologies have been developed. Despite this popularity, only a very few methods were created and even these struggle with one or more drawbacks. We reckon that the practical aspect of working with ontologies is the problem that holds this technology back. In this paper we briefly describe some of existing methods for evaluation and visualization of ontologies which possibly have potential pull the problem off.

**Keywords**—ontology, ontology visualization, ontology evaluation.

## I. INTRODUCTION

Web content is growing fast and it is becoming rather complicated to find relevant information. Semantic web is an initiative to make web more usable, acceptable and comprehensible for both users and machines. In order to make semantic web practical is necessary to provide new functionalities and additional information about web content. From this perspective, ontologies can be seen as repositories of such information captured by means of standardized semantic web technologies. Ontology is a formal, unambiguous delimitation of shared terms [1]; i.e., ontology provides a shared dictionary, which describes the chosen domain, the types of objects and terms, their attributes and relations among them. Categories (or concepts), their attributes and relations among them create the core of an ontology, often called TBox.

There are several semantic web engines that can be used to find ontologies. These are mostly based on keyword searching; therefore they may not always provide relevant results, since keywords are too ambiguous for describing ontologies. After the list of possibly relevant documents is obtained from such an engine, the user needs to decide which of them will be used. As ontologies are not in a human readable form, it is necessary to use one of the visualization tools to explore these results. There are several techniques that help to navigate in ontologies and thus can be used to discover objects, relations and properties in ontology. Based on this information the user is able to make a decision whether a particular model is appropriate or not to his or her domain.

Techniques allowing the user to navigate in ontology and explore concepts belong among ontology visualization tools. Tools from this group contain many different ways how to display objects, properties and instances from ontology. In the

field of ontology visualization, there are several techniques borrowed from other contexts, such as graph topology or file system visualization, that could be adapted to display ontologies.

In spite of wide variety of different approaches, none of them manages to cope with complexity of visualization problem. All these tools struggle with one or more aspects of ontology visualization. We do not attempt to replace these approaches but complete the offer that already exists and therefore fill in the gap we have spotted on the field of ontology visualization.

The most common problem (which many approaches are not able to cope with) is information overflow issue. This issue emerges when a lot of information needs to be showed in a relatively small area. Furthermore, not all data included in the ontology are necessary (for visualization) in order to understand what the ontology is about or whether it fits a particular topic. Topical understanding of ontologies, though, is a part of another branch of research – that is ontology evaluation.

It is believed that ontology evaluation and visualization are sibling issues; they are both parts of a bigger problem. Our aim is to provide a tool capable of solving both sub-tasks (evaluation and visualization), which obviously requires better understanding of the complexity and nature of each of these problems.

## II. ONTOLOGY

### A. Definition

One of the possible definitions of ontology: Ontology is a formal, unambiguous delimitation of shared terms. It means that it provides shared dictionary which describes the chosen domain, the types of objects and terms, their attributes and relations among them. Categories, their attributes and relations among them create the ground of ontology. It is necessary to define them; it means to note the category, determine its attributes as well as its relation to other categories. After the definition of categories it is possible to assign them some instances, which will represent the objects of these categories. For the reason that the relations among objects are known, it is possible to derive some new facts [2].

### B. Ontology and Semantic Web

One of the fundamental problems in semantic web is that one and the same term can be labeled by different codes.

Agent in such case must have a possibility to differentiate, what meaning has information, for given data source it can get in touch with.

It is just ontology that was designed for this problem solving. The term ontology represents the summary of formally defined relations among the objects. In ontology so called derivational rules are used. They serve for derivation of other relations among entities.

Ontology increases the functionality of web. In this way that it allows to specify the web retrieval. With help of ontology, program should be able to display, only relevant pages for current query. This is the most direct utilization of ontology.

There exist also more sophisticated accesses, for example:

- the using of association of knowledge structure
- derivation of ontology rules

Such a use could lead to the complete automatic retrieval of all the relevant references.

### III. ONTOLOGY EVALUATION

Widely accepted definition of ontology evaluation is: *Ontology evaluation is the problem of assessing a given ontology from the point of view of a particular criterion of application, typically in order to determine which of several ontologies would best suit a particular purpose* [3]. In the recent years the focus has shifted from data towards knowledge processing. As a result, the basic processing unit is less and less atomic piece of data. Hence, the ability to evaluate and compare the ideas within the area makes this scope interesting.

Ontology evaluation is therefore important problem that need to be solved in order to make ontologies more practical and usable. Users must have the way of deciding which ontology is the most appropriate for certain application eventually which sub set of ontologies might be used.

Techniques developed in the area of evaluation might be also interesting for ontology developers who need to rate results in process of development and according results alter or redefined goals.

Since in ontology is a fairly complex structure, it is often practical to on the evaluation of different levels of the ontology separately rather than to do it complexly. According [3] ontology evaluation can be performed on the following levels:

*Lexical, vocabulary, or data layer.* Here the focus is on which concepts and instances have been included in the ontology, and the vocabulary used to represent or identify these concepts. Evaluation on this level tends to involve comparisons with various sources of data concerning the problem domain, as well as techniques such as string similarity measures or so.

*Hierarchy or taxonomy.* Ontology typically includes objects and hierarchical relations among them. Obviously, some ontology languages provide broader range of relations that can be defined and therefore enhance information value of such a data.

*Other semantic relations.* The ontology may contain other relations besides is-a, and these relations may be evaluated separately. This typically includes measures such as precision

and recall.

*Context or application level.* An ontology may be part of a larger collection of ontologies, and may reference or be referenced by various definitions in these other ontologies. In this case it may be important to take this context into account when evaluating it. Another form of context is the application where the ontology is to be used; evaluation looks at how the results of the application are affected by the use of the ontology.

*Syntactic level.* Evaluation on this level may be applied on ontologies that have been mostly constructed manually. The ontology is usually described in a particular formal language and must match the syntactic requirements of that language. Various other syntactic considerations, such as the presence of natural-language documentation, avoiding loops between definitions, etc., may also be considered.

*Structure, architecture, design.* This is primarily of interest in manually constructed ontologies. We want the ontology to meet certain pre-defined design principles or criteria; structural concerns involve the organization of the ontology and its suitability for further development. This sort of evaluation usually proceeds entirely manually.

### IV. ONTOLOGY VISUALIZATION

Ontology is often depicted as a hierarchy of concepts. Sometimes, such a hierarchy is enriched with role relations among concepts and each concept has various attributes related to it. Further each concept most likely has instances attached to it, which could range from one or two to thousands. Therefore, it is not trivial to create a visualization that would effectively display all information and allow user to easily perform various operations on the ontology. The methods can be grouped according to different characteristics of the presentation, interaction technique, and functionality into several groups: indented list, node-link and tree, zoomable, space-filling, focus + context or distortion, 3D information landscapes. [4]

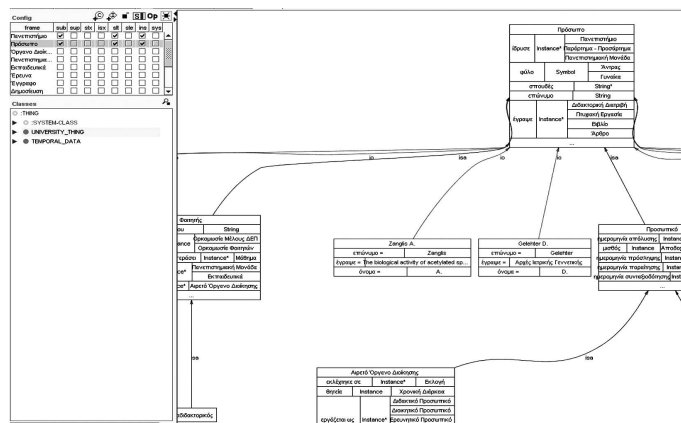


Figure 1: OntoViz window screenshot

Even though there were some tools developed but all of them struggle to cope with one or more issues. Here we'd like to discuss two of them regarding highlight and distinguish improvements that we have made. First of them is OntoViz [5] that is visualization plug-in for Protégé. OntoViz creates 2D graph with the capability for each class to present its properties and inheritance as it is shown on Fig.1. Many users

claim lack of interaction and problems with navigation. Furthermore, they found presentation “poor” and commented on the lack of a search tool. [6]

Other visualization tool which represents class hierarchy as a set of concentric circles is CropCircles [7]. Child nodes are placed as concentric circles in the parent circle to its parents as it is displayed on Fig.2. The user may click on a circle to highlight it and see a list of its immediate children on a selection pane. This can let the user drill down the class hierarchy level-by-level. Problem of CropCircles method is poor space filling that is caused by top-down layout. The problem is becoming even more important for big ontologies.

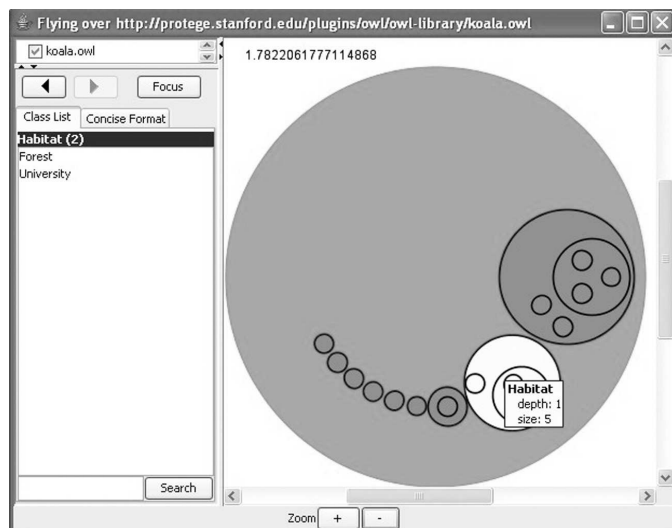


Figure 2: The CropCircles visualization in Swoop.

Our goal is not in replacing existing techniques but filling in the gap we have spotted in ontology visualization field.

## V. CONCLUSION

In the paper we have presented that the problem of comprehending an ontology consists of two sub-tasks, namely, ontology evaluation and ontology visualization. Both are equally important and together yield an ability of making well-informed decisions; they provide a new point of view on ontologies than has been available so far. A significant number of visualization and evaluation tools struggle with resolution issues occurring when a large amount of information needs to be displayed. Rather than showing all topological classes, properties and instances we preferred data filtering, results of which are summarized and presented to the user. This is based on assumption that not all concepts existing in an ontology are necessary to be visually shown in order to understand the domain of a particular document.

Reducing the information overload is often cited as the premise for work on supporting the retrieval of relevant documents. Though finding relevant documents is only half of the task. The rest is mostly on interpretation and evaluation. And those are issues we would like to deal with.

## ACKNOWLEDGMENT

The work presented in this paper was supported by the Slovak Grant Agency of Ministry of Education and Academy of Science of the Slovak Republic within the 1/4074/07

project ”Methods for annotation, search, creation, and accessing knowledge employing metadata for semantic description of knowledge”.

## REFERENCES

- [1] Gruber, T.R.: Toward principles for the design of ontologies used for knowledge sharing.
- [2] M. C. Daconta, Leo J. Obrst, Kevin T. Smith, “The Semantic Web,” in Indiana: Indiana Polis, 2003, pp. 57–237.
- [3] Brank, J., Grobelnik, M., Mladenic, D.: A Survey of Ontology Evaluation Techniques, Department of Knowledge Technologies Jozef Stefan Institute, (2005)
- [4] Katifori, A., Halatsis, C., Lepouras, G., Vassilakis, C., and Giannopoulou, E. 2007. Ontology visualization methods—A survey. *ACM Comput. Surv.* 39, 4, Article 10 (October 2007), 43 pages DOI = 10.1145/1287620.1287621 <http://doi.acm.org/10.1145/1287620.1287621>.
- [5] Sintek, M. 2003. Ontoviz tab: Visualizing Protégé ontologies, <http://protege.stanford.edu/plugins/ontoviz/ontoviz.html>.
- [6] Akrivi, K. Elena, T. Constantin, H. Georgios, L. Costas, V.: A Comparative Study of Four Ontology Visualization Techniques in Protege: Experiment Setup and Preliminary Results (2006)
- [7] Parsia, B., Wang, T., Golbeck, J. Visualizing Web Ontologies with CropCircles

# On the semantic correspondence of B and BPA specifications

<sup>1</sup>Marek Výrost, <sup>2</sup>Attila N.Kovács

<sup>1,2</sup>Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

<sup>1</sup>marvy@reacode.com, <sup>2</sup>Attila.N.Kovacs@tuke.sk

**Abstract**—Our work presents a link between two different types of specification of behavior of a computing object: process algebraic specification based on bisimulation equivalence and B-method based on weakest-precondition semantics. We deal with correspondence at the level of semantics and show how to find appropriate abstract machine for any definable basic process.

**Keywords**—correspondence, B-method, basic process algebra, semantics

## I. PREREQUISITES

The kind reader is expected to possess basic knowledge of B-Method[1] and basic process algebra in the style of algebra of communicating processes [2], [3].

## II. TWO APPROACHES TO SPECIFICATION OF BEHAVIOR

Behavior of an object of the real world which is realized by a program (or computation) is in general specifiable in several ways. In this paper we will be dealing with two of them in particular.

The first approach is focused on interactions in which the object is able to be involved. In this case we specify behavior from an observational point of view, without describing the object's internal structure. Behavior of the object is specified by sequences of atomic behavioral elements. These atomic behavioral elements are usually called actions.

The other approach to specification of an object's behavior is based on a description of object's properties and their changes. We refer to actual set of properties which are attributed to the given object as the *state* of the object. Behavior of the object is then observed as sequence of changes of its state.

Both approaches mentioned above may be illustrated by an elevator in a building with three floors. Assuming that the elevator starts at the lowest floor, we can describe its behavior using the first approach like this: an elevator can initially only go up. Then it can go down and behave like an elevator which was not moved yet; or it can go up and then go up or down... and so on. If we want to apply the second approach, we have to specify some properties of the elevator and then define the means by which these properties are changed. Our choice of property is obvious - we notice that elevator is at certain floor, so the changing property will be the number of floor the elevator is at.

## III. BPA $\circ$ B

Let us denote a set of all labelled transition systems by  $\mathcal{U}_{LTS}$ , set of all BPA process expressions by  $\mathcal{U}_{\mathcal{P}}$  and a set of all abstract machines by  $\mathcal{U}_{\mathcal{B}}$ .

Generalized relation of operation semantics for process expressions will be denoted as  $\rightarrow \subseteq \mathcal{U}_{\mathcal{P}} \times \mathcal{U}_{LTS}$ .

For every abstract machine  $M \in \mathcal{U}_{\mathcal{B}}$ , we can find its labelled transition system  $LTS_M$ . Relation between the set of abstract machines and the set of labelled transition systems will be denoted as  $\dashrightarrow \subseteq \mathcal{U}_{\mathcal{B}} \times \mathcal{U}_{LTS}$ .

*Definition 1:* BPA process  $P$  is represented by abstract machine  $M$  (denoted  $P \circ M$ ), iff

$$P \rightarrow LTS_P \wedge \exists LTS_M (M \dashrightarrow LTS_M \wedge LTS_P \approx LTS_M).$$

This relationship is depicted in figure 1.

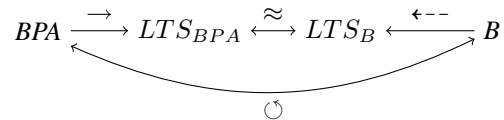


Fig. 1. Relationship between BPA and B specifications

This definition says that a process (which may be a solution of a recursive BPA specification) and abstract machine may be considered behaviorally equivalent, if its labelled transition system is related to labelled transition system generated by abstract machine. This relation is known as bisimulation.

The relationship between BPA and B specifications can be illustrated by our example with elevator.

## IV. AN ELEVATOR IN BPA

BPA specification of a process modeling behavior of an elevator in a building with 3 floors is

$$\begin{aligned} \underline{OnlyUp} &= up.UpAndDown \\ UpAndDown &= down.OnlyUp + up.OnlyDown \\ \underline{OnlyDown} &= down.UpAndDown \end{aligned} \quad (1)$$

An operational semantics of this specification is defined by action relations  $\xrightarrow{up}, \xrightarrow{down}$ :

$$\begin{aligned} \xrightarrow{up} &= \{(OnlyUp, UpAndDown), (UpAndDown, OnlyDown)\} \\ \xrightarrow{down} &= \{(UpAndDown, OnlyUp), (OnlyDown, UpAndDown)\} \end{aligned}$$

Labelled transition system of the elevator,

$$LTS_V = \{(OnlyUp, UpAndDown, OnlyDown), \{\xrightarrow{up}, \xrightarrow{down}\}\}$$

is depicted in figure 2.

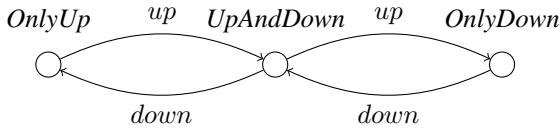


Fig. 2. Labelled transition system of an elevator with three floors

## V. FROM PROCESSES TO STATES

Performing an assignment of a change of properties to each of object's actions leads us from specification of behavior as sequences of actions to specification of behavior based on states.

In mathematical model, the state of an object is represented by a variable. The variable takes a distinguished value from a set of admissible values. The value of a variable thus represents actual property of the object and a set of all admissible values of the variable represents all of its observable properties.

*Definition 2:* Let  $(\mathcal{P}, \{r_i^i \mid r_i^i \subseteq \mathcal{P} \times \mathcal{P} \wedge i \in I\})^1$  be a labelled transition system of a process  $P \in \mathcal{P}$  and let  $\mathcal{S}_p$  be a set of states. Representation  $\mathfrak{R}_P$  of the process  $P$  is a function  $\mathfrak{R}_P : \mathcal{P} \rightarrow \mathcal{S}_p$ .

Speaking about the elevator, we have chosen the number of floor the elevator is at as its property. Let us now assign a predicate to every process of specification (1) in such a way, that in every moment only one of these predicates holds. Each predicate will express certain observation, in our case the number of floor.

Representation  $\mathfrak{R}_{OnlyUp}$  is then defined as follows.

$$\begin{aligned} OnlyUp &\rightarrow \langle floor = 1 \rangle \\ UpAndDown &\rightarrow \langle floor = 2 \rangle \\ OnlyDown &\rightarrow \langle floor = 3 \rangle \end{aligned}$$

Substitution of processes in the definition of action relations  $\xrightarrow{up}, \xrightarrow{down}$  for these predicates gives us following the form of action relations:

$$\xrightarrow{up} = \{ \langle floor = 1 \rangle, \langle floor = 2 \rangle, \langle floor = 2 \rangle, \langle floor = 3 \rangle \} \quad (2)$$

$$\xrightarrow{down} = \{ \langle floor = 2 \rangle, \langle floor = 1 \rangle, \langle floor = 3 \rangle, \langle floor = 2 \rangle \} \quad (3)$$

## VI. MODEL OF AN ABSTRACT MACHINE

Set of variables of abstract machine  $M$  will be denoted as  $\mathcal{V}^M$ . Set of operations of abstract machine  $M$  will be denoted as  $\mathcal{O}^M$ . Invariant of abstract machine  $M$  will be denoted as  $\mathcal{I}^M$ . Initialization substitution of machine  $M$  will be denoted as  $i^M$ . Substitution of operation of  $M$  will be denoted as  $s_{op}$ .

## VII. A FEW CONCEPTS FROM THEORY OF ABSTRACT MACHINES

Body of operation of abstract machine is a generalized substitution. Generalized substitution of an operation of abstract machine specifies not only the new state after its execution, but also context within which the operation should be executed, if the invariant of the machine is to be preserved.

<sup>1</sup>In this paper,  $I$  is considered to be an index set.

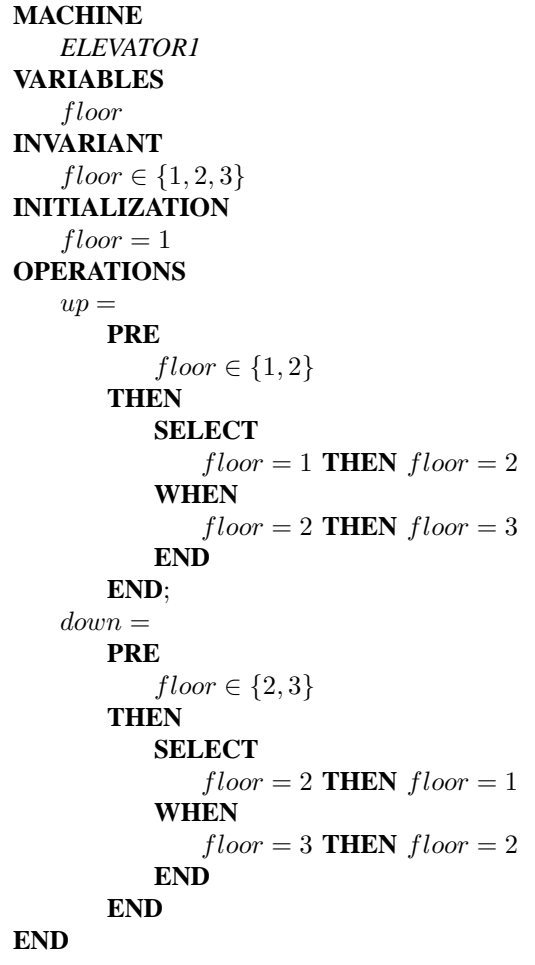


Fig. 3. Abstract machine for an elevator with three floors (version 1)

If an operation is executed in a state which is within its valid context, we say that the substitution (of operation) terminates. Predicate, which holds in the case where substitution  $\mathfrak{s}$  working with variable  $x$  terminates, is defined as

$$trm_x(\mathfrak{s}) = [\mathfrak{s}](x = x)$$

Operations of an abstract machine do change its state. Semantics of an operation may be therefore defined as a relation between old and new value of variables of abstract machine. This relation is expressed by so called before-after predicate. The new value of variable after an execution of an operation is written as primed:  $variable' = f(variable)$ . Before-after predicate of generalized substitution  $\mathfrak{s}$  working with variable  $x$  is defined as

$$prd_x(\mathfrak{s}) = \neg[\mathfrak{s}](x' \neq x).$$

A set of admissible values should be specified in an invariant of abstract machine for every variable. Membership of variable's value in this set must stay invariant with respect to all operations within machine. We can extend the concept of termination of generalized substitution  $\mathfrak{s}$  working with variable  $x$ , whose values are members of set  $s$  and also extend its before-after predicate to sets of values. This extension leads to a set of preconditions

$$pre(\mathfrak{s}) = \{x \mid x \in s \wedge trm(\mathfrak{s})\}$$



and a relation expressing the dynamics of substitution

$$rel(\mathfrak{s}) = \{(x, x') \mid x, x' \in s \times s \wedge prd_x(\mathfrak{s})\},$$

defined as a set of pairs of values of variable  $x$  before and after the execution of substitution  $\mathfrak{s}$ .

The set  $pre(\mathfrak{s})$  and relation  $rel(\mathfrak{s})$  completely characterize the generalized substitution  $\mathfrak{s}$ . Every generalized substitution  $\mathfrak{s}$  can be rewritten in the form

$$\mathfrak{s} = trm(\mathfrak{s}) \mid @x'(prd_x(\mathfrak{s}) \Rightarrow x := x')$$

For any abstract machine having a predicate in the form  $x \in s$  for some set  $s$  as a part of invariant and any generalized substitution  $\mathfrak{s}$  being a body of one of machine's operation, following holds:

$$\mathfrak{s} = x \in pre(\mathfrak{s}) \mid @x'((x, x') \in rel(\mathfrak{s}) \Rightarrow x := x').$$

Let  $\mathfrak{s}$  be a generalized substitution working with variable  $x$  of a abstract machine, whose invariant contains predicate in the form  $x \in s$ . Following relationship between set  $pre(\mathfrak{s})$  and relation  $rel(\mathfrak{s})$  can be found:

$$\overline{pre(\mathfrak{s})} \times s \subseteq rel(\mathfrak{s}).$$

*Lemma 1:* Let  $s$  be a set,  $p \subseteq s$ ,  $r \subseteq s \times s$ ,  $\bar{p} \times s \subseteq r$  and let  $\mathfrak{s}_r$  be a generalized substitution  $\mathfrak{s}_r = x \in p \mid @x'((x, x') \in r \Rightarrow x := x')$ . Then

$$(pre(\mathfrak{s}_r) = p) \wedge (rel(\mathfrak{s}_r) = r).$$

Relation  $r$  from lemma 1 is called *relation corresponding to generalized substitution  $\mathfrak{s}_r$* .

## VIII. ABSTRACT MACHINE OF A BPA PROCESS

Action relations are defined by means of pairs of processes - each pair consists a process before and process after execution of concrete action. A substitution of states of abstract machine for processes in an action relation converts the action relation to a relation describing states before and after execution of corresponding action.

For example, action relations (2),(3) are in the form

$$\{(floor, floor')\} \subseteq \{1, 2, 3\} \times \{1, 2, 3\}.$$

It is not hard to see similarity with set of before-after predicates for generalized substitution working with variable  $floor$  on the set of values  $\{1, 2, 3\}$ . This property holds universally without difference for any chosen structure of state of abstract machine. This brings us to following theorem.

*Theorem 1:* For every action relation  $\xrightarrow{a}$  of action  $a$  on a set of states  $\mathcal{S}_a$  it is possible to find its corresponding generalized substitution  $\mathfrak{s}_a$ .

**Proof**

Let us define relation

$$r_a = \xrightarrow{a} \cup (\mathcal{S}_a - dom(\xrightarrow{a})) \times \mathcal{S}_a.$$

Then

$$\mathfrak{s}_a = x \in dom(\xrightarrow{a}) \mid @x'((x, x') \in r_a \Rightarrow x := x')$$

is a generalized substitution with property

$$(pre(\mathfrak{s}_a) = dom(\xrightarrow{a})) \wedge (rel(\mathfrak{s}_a) = r_a)$$

as a consequence of lemma 1 with  $p = dom(\xrightarrow{a})$ ,  $r = r_a$ .  $\square$

*Theorem 2:* Let  $(\mathcal{P}, \{\xrightarrow{r_i} \mid \xrightarrow{r_i} \subseteq \mathcal{P} \times \mathcal{P} \wedge i \in I\})$  be a labelled transition system of process  $P \in \mathcal{P}$  and let  $\mathfrak{R}_P$  be a representation of process  $P$  with set of states  $\mathcal{S}_P$ . Then for every action relation  $\xrightarrow{r_i}$  for  $i \in I$  there is an operation  $op(\xrightarrow{r_i})$  with generalized substitution  $\mathfrak{s}_{op(\xrightarrow{r_i})}$  and there is an abstract machine  $M$  with:

- set of variables  $\mathcal{V}^M = \{state\}$
- invariant<sup>2</sup>

$$\mathcal{I}^M = \bigcup_{i \in I} (dom(pre(\mathfrak{s}_{op_i^M})) \cup ran(dom(pre(\mathfrak{s}_{op_i^M})) \triangleleft rel(\mathfrak{s}_{op_i^M}))) \cup \{\mathfrak{R}_P(P)\}$$

- initialization  $i^M = [state := \mathfrak{R}_P(P)]$  and
- operations  $\mathcal{O}^M = \{op_i^M \mid i \in I \wedge op_i^M = op(\xrightarrow{r_i})\}$

which will satisfy all proof obligations.

**Proof**

Initialization of abstract machine  $M$  must establish invariant:

$$\mathfrak{R}_P(P) \in \mathcal{I}^M$$

and every operation  $op_i^M$  with generalized substitution  $\mathfrak{s}_{op_i^M}$  must preserve it (if an operation is executed from state in which invariant holds, then the new state after its execution must satisfy this invariant as well):<sup>3</sup>

$$\begin{aligned} dom(pre(\mathfrak{s}_{op_i^M})) &\subseteq \mathcal{I}^M \\ ran(pre(\mathfrak{s}_{op_i^M}) \triangleleft rel(\mathfrak{s}_{op_i^M})) &\subseteq \mathcal{I}^M \end{aligned}$$

Satisfaction of all conditions mentioned previously is clear from the construction of invariant  $\mathcal{I}^M$  of abstract machine  $M$ .

Direct translation of action relations  $\xrightarrow{up}$  and  $\xrightarrow{down}$  produces B-machine depicted in figure 3.  $\square$

## IX. SUBSTITUTIONS AS COMBINATIONS OF PRIMITIVE RELATIONS

Abstract machine from figure 3 has a disadvantage in enumeration of all elements of action relations. Just imagine what would it take to specify an elevator in a building with 150 floors this way. In such cases the presented approach of representation of action relations is inappropriate.

In practice, action relations can be defined by reusing existing relations, like arithmetic operations on the set of natural numbers. For example, take action  $up$  and consider relation

$$\begin{aligned} +_1 &= \{(x, y) \mid (x, y) \in \mathbb{N} \times \mathbb{N} \wedge y = x + 1\} \\ &= \{(1, 2), (2, 3), (3, 4), \dots\} \end{aligned}$$

The relationship between values of its elements is exactly the one we are looking for. However, we do need to specify its domain restriction  $\{1, 2\} \triangleleft +_1$  to ensure the preservation of invariant. Now the same problem of enumeration of all elements applies to the definition of restriction. In this case there is other standard relation on natural numbers, which

<sup>2</sup>Every predicate  $P(x)$  with free variable  $x$  taking values from set  $S$  can be interpreted as a subset of  $S$ .

<sup>3</sup> $\triangleleft$  denotes domain restriction. Let  $r \subseteq R \times R$  be a relation and  $s \subseteq R$  a set. Then  $s \triangleleft r = \{(a, b) \mid (a, b) \in r \wedge a \in s\}$ .

allows us to define the needed set of elements:

$$\begin{aligned} < &= \{(x, y) \mid (x, y) \in \mathbb{N} \times \mathbb{N} \wedge \exists z \in \mathbb{N}(y = x + z)\} \\ &= \{(1, 2), (1, 3), (2, 3), (2, 4), \dots\} \\ pre(\overset{up}{\rightarrow}) &= \{x \mid x \in \mathbb{N} \wedge (x, 3) \in <\} = \{1, 2\} \end{aligned}$$

Combination of the two relations gives us following representation of action relation  $up^4$ :

$$\begin{aligned} \overset{up}{\rightarrow} &= \{x \mid (x, 3) \in <\} \triangleleft +1 \\ &= \{(x, x') \mid (x, x') \in \mathbb{N} \times \mathbb{N} \wedge x < 3 \wedge x' = x + 1\} \end{aligned}$$

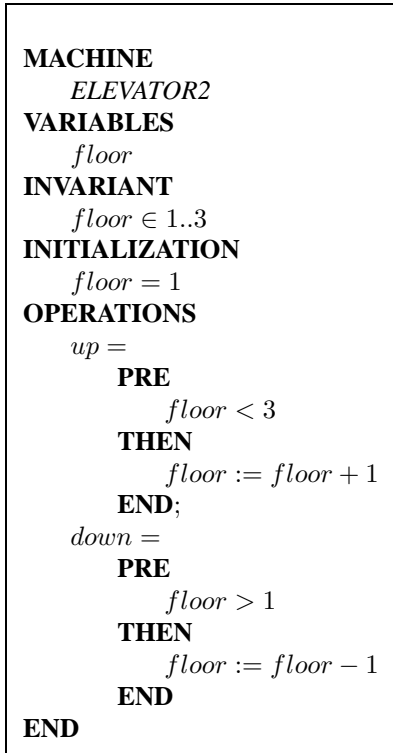


Fig. 4. Abstract machine for an elevator with three floors (version 2)

Resulting abstract machine is depicted in figure 4.

### X. RELATED WORK

The relationship between basic process algebra and B-Method has not been studied extensively yet. However, basic process algebra is a special instance of a family of formalisms named process algebras and one of the other members of this family is CSP.

The sharp distinction between these formalisms lies in their semantics: the central notion in BPA is bisimilarity, while in CSP, the semantics is defined by traces. BPA follows traditional theory of universal algebra. It can be extended with more complex operators on processes, like interleaving, concurrency and specification of time [2].

Helene Treharne and Steve Schneider [4] suggested an approach where operations of B-machines are called from CSP parameterized mutual recursion process. Theory of this approach can be found in [5] and [6], its practical application in [7].

The other approach, taken by Michael Butler, suggests incorporating a CSP specification within the definition of B-Machine [8].

<sup>4</sup>Variables  $x, x'$  are in this context universally quantified, and thus alpha-convertible with variables  $floor, floor'$ .

### XI. CONCLUSION

We have proposed an approach for generation of abstract machines for processes which are solutions of basic process algebra specifications. Our approach differs from the work of other authors in several aspects. We deal with basic process algebra, not CSP. Any extension of basic process algebra is a potentially subject to our integration.

Our approach does not require any syntactic extensions of abstract machine specification, because it works at the level of semantics.

### REFERENCES

- [1] J.-R. Abrial, *The B-book: assigning programs to meanings*. New York, NY, USA: Cambridge University Press, 1996.
- [2] J. C. M. Baeten, C. A. Middelburg, and K. Middelburg, *Process Algebra with Timing*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2002.
- [3] J. C. M. Baeten, Ed., *Applications of process algebra*. New York, NY, USA: Cambridge University Press, 1990.
- [4] H. Treharne and S. Schneider, "Using a process algebra to control B OPERATIONS," in *IFM'99 1st International Conference on Integrated Formal Methods*. York: Springer-Verlag, 1999, pp. 437–457.
- [5] N. Evans and H. Treharne, "Linking semantic models to support csp || b consistency checking," *Electr. Notes Theor. Comput. Sci.*, vol. 145, pp. 201–217, 2006.
- [6] S. Schneider and H. Treharne, "Csp theorems for communicating b machines," *Formal Asp. Comput.*, vol. 17, no. 4, pp. 390–422, 2005.
- [7] N. Evans and H. Treharne, "Investigating a file transfer protocol using csp and b," *Software and System Modeling*, vol. 4, no. 3, pp. 258–276, 2005.
- [8] M. Butler, "csp2b: A practical approach to combining csp and b," *and B. Formal Aspects of Computing*, vol. 12, p. 2000, 1999.

# Diagram of security

*Marek VYSOKÝ*

Department of Computers and Informatics, FEI TU of Košice, Slovak Republic

mvysoky@lundegaard.sk

**Abstract**— this paper describes how to modeling security in real network systems and how to get tools for determine security risk and treatments from attackers and users of system from inside system environment and outside system environment.

**Keywords**—Attack, Deploy diagram UML 2.x, Network systems, Management of risk, Security diagram, Security Modeling, Security Incident, Treatments,

## I. INTRODUCTION

When we are modeling of informatics systems we are using various models such as UML models, network diagrams, mathematical models, simulation models etc. Each phase of modeling brings level of knowledge for informatics systems. Obtained models allow define behavior or properties of system. I describe possibilities of security modeling and describe tools for determining risk of threats from outside or inside environment. I describe utilities for analysis existing or creating systems, which result determines level of security of modeling systems. This work is included in project Knowledge-Based Software Life Cycle and Architectures [9]. One of the goals of the project is software system architecture analysis suitable for knowledge integration about system and application domain.

## II. MODELING OF SECURITY

Modeling of security requires mathematical and graphical equipment. Security of informatics systems is wide area. I'm orienting to modeling of security in organization and network models. There are many IT elements in organization such as workstation, servers, firewalls, network disks and many other network elements. Modeling of security in this level required knowledge of interconnections of several elements and their security.

Diagram of security contains elements:

1. Entities – elementary elements of network such as workstation, firewalls, servers, etc
2. Data source – data storage such as databases, file systems , etc
3. Users – users from inside environment and from outside environment
4. Connections – connection between elements in diagram

Entities are elementary elements of network. Security definitions of the entities are:

1. Software security – it is coefficient, which designates security level of used software and operating system, for example security software used in the workstation or server in light on potential attack and data security.

Determination of this coefficient requires knowledge about used software and operation systems in given entity. On the basis of knowledge about details of software is possible determine software security coefficient for all entity.

2. Number of incidents – is number, which designates number of known security incidents. Source for determines this number may be security log. Number may be computed as weighted average of all incidents in security log. Weight of incident determines more dangerous incidents and increase security risk of incident.

Intersection between entities in diagram defines various security parameters. Intersection defines security and immunity in face of attacks or unintended violation of security or assignment sensitive data.

Therefore we define parameters:

1. Cryptography level – it is coefficient, which determine level of used cryptography. We are able to create table of used cryptography and for each one determine coefficient with the view to security of transmitted data and level of cryptography algorithm and keys.
2. Security of transmitting data – is coefficient, which determines sensitivity of transmitting data by this channel with the view to possibility to their abuse.

I want to separate out data source from other entities in diagram of security. Data source representing data storage. In much case is data storage on the server's facilities and it could it be included in set of entities, but for analysis risk and possible attack it is conclusive find path special to data sources and risk of their abuse.

In this case determine those parameters:

1. Data sensitivity - it is coefficient, which determine sensitivity of data. Determine scale of their confidentiality and potential threat in the case their exploitation.
2. Number of incidents – is number, which designates number of known security incidents. Source for determines this number may be security log. Number may be computed as weighted average of all incidents in security log. Weight of incident determines more dangerous incidents and increase security risk of incident.

Significant factor for modeling of security is human. Human (User) use to entities in network by using computers facilities (work station, PDA ...). Users, which are entering to systems or network have variety of computer education and have various abilities.

Therefore we can define those parameters:

1. Knowledge – evaluation parameter, which determine IT ability of user of system. Ability of administrator, which manages entities of the system are much higher than ability of maintaining worker, which uses systems for evidence.
2. Power of facilities –this parameter determines power of technical facilities, which is able to use to attack. Surely, salesman of organization use notebook, which computing power is insignificant compare to organized group with high-tech facilities, which want to attack to network towards espionage.
3. Numbers of access – this parameter determine number of access to network or organization structure. It is possible to get it from security logs and make actual. Users of system, who have this number higher, are potentially more risky than other.

Vector of access – each users who use entity have assign level of access. Vector of access define accessibility to elements through to connection in the model. Vector of access may be trivial (users have or not have access), or more extensive (expand access to read, write, delete, change...).

### III. MATHEMATICAL REPRESENTATION OF DIAGRAM OF SECURITY

Mathematical representation of diagram of security is coming from basic definition.

Diagram of security is  $D_b = (U, E, DS, \alpha, \beta)$

$U$  - is finite set of users from inside environment and outside environment.

Attributes of this element are:

- Knowledge  $Uk$
- Power of facilities  $Up$
- Number of access  $Una$

$E$  - is finite set of entities (workstation, firewalls, servers...)

Attributes of this element are:

1. Software security  $Ess$
2. Number of incidents  $Eni$

$DS$  - is finite set of data sources (databases, data storage, file systems ...)

Attributes of this element are:

1. Data sensitivity  $DSds$
2. Number of incidents  $DSni$

$\alpha : (ExE) \rightarrow 1 | 0$  - is projection defines relation between entities. Connection between entities exists if result is 1. It means for  $(x, y) \in E^2$  applies  $\alpha(x, y) = z$

where  $z \in (0,1)$ . Practically it means, that if result of projection is 1, exists intersection between entity  $x$  and entity  $y$ . [1]

Attributes of intersection edge:

1. Cryptography level  $TC$
2. Security of transmitting data  $DSec$

$\beta : (ExDS) \rightarrow 1 | 0$  - is projection defines relation between entities and data sources. Connection between entity and data

source exists if result is 1. It is means for  $(x \in E, y \in DS)$  apply  $\beta(x, y) = z$  where  $z \in (0,1)$ . Practically it means, that if result of projection is 1, exists intersection between entity  $x$  and data source  $y$

Attributes intersection edge

1. Cryptography level  $TC$
2. Security of transmitting data  $DSec$

$\omega : Ux(ExE) \rightarrow (0,1)$  - is projection, which determines privileges of user to access to entity. When user has access then result is 1 else 0. It is means for:

$(u \in U, x \in E, y \in E, \alpha(x, y) = 1)$  apply

$\omega(u, (x, y)) = z$  where  $z \in (0,1)$ .

$\delta : Ux(ExDS) \rightarrow (0,1)$  - is projection, which determines privileges of user to access to data source. When user has access then result is 1 else 0. It is means for:

$(u \in U, x \in E, y \in DS, \beta(x, y) = 1)$  apply

$\delta(u, (x, y)) = z$  where  $z \in (0,1)$ . Extension of definition of access is modification of result of operation to:

$z \in (read, write, delete, ...)$

Vector of access between entities:

$v_e(x, y) = (\omega(u1, (x, y)), \omega(u2, (x, y)) \dots \omega(un, (x, y)))$

where  $un \in U, x \in E, y \in E$

Vector of access between entities and data sources:

$v_{eds}(x, y) = (\omega(u1, (x, y)), \omega(u2, (x, y)) \dots \omega(un, (x, y)))$

where  $un \in U, x \in E, y \in DS$

Adjacency matrix:

$k_i, k_j \in (E \cup D)$

$a_{ij}$  - elements of matrix

$a_{ij} = \begin{cases} 1 & \text{if } k_i \text{ is starting vertex and } k_j \text{ is ending vertex, } \beta(k_i, k_j) = 1 \text{ or } \alpha(k_i, k_j) = 1 \\ 0 & \text{in other cases} \end{cases}$

Matrix of user access:

User  $u \in U, k_i, k_j \in (E \cup D)$

$p_{ij} = \begin{cases} \omega(u, (k_i, k_j)) = 1 & k_i \in E \text{ is starting vertex and } k_j \in E \text{ is ending vertex.} \\ \delta(u, (k_i, k_j)) = 1 & k_i \in E \text{ is starting vertex and } k_j \in DS \text{ is ending vertex} \\ 0 & \text{in other cases} \end{cases}$

IV. GRAPHICAL REPRESENTATION OF DIAGRAM OF SECURITY

Graphical representation is coming out from mathematical representation.

User from set of users  $U$  represents by symbol



Name of user is displaying as label in the symbol.

Entity from set of entities  $E$  represents by symbol



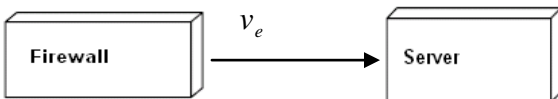
Name of entity is displayed as label in the symbol.

Data source from set of data sources  $DS$  represents by symbol.



Name of entity is displayed as label in the symbol.

Projection  $\alpha$  -intersection between entities represents by oriented edge. Vector of access  $v_e$  is displayed as label of the edge.

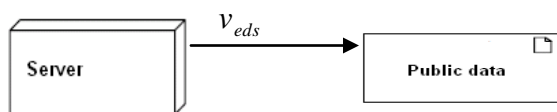


User has direct access to entity. This is main point for user to entry system or network.



Projection  $\delta$  -intersection between entity and data source represents by oriented edge and label for edge is defined as

vector of access  $v_{eds}$



V. ANALYSIS DIAGRAM OF SECURITY

A. Weak places in diagram of security

When we want to search for weak places in diagram of security we can use algorithm to find shortest path in graph. Acceptable algorithm can be Floyd – Warshall algorithm. Floyd–Warshall algorithm – sometimes known as the WFI Algorithm or Roy–Floyd algorithm, since Bernard Roy described this algorithm in 1959 is a graph analysis algorithm for finding shortest paths in a weighted, directed graph. A single execution of the algorithm will find the shortest paths between all pairs of vertices. In our case we interest in shortest path from user to data source, or to entity. Fundamentals of algorithm:

Consider a graph  $G$  with vertexes  $V$  each numbered 1 to  $N$ . Further consider a function  $shortestPath(i, j, k)$ ,

that returns the shortest possible path from  $i$  to  $j$  using only vertices 1 to  $k$  as intermediate points along the way.

Now, given this function, our goal is to find the shortest path from each  $i$  to each  $j$  using only nodes 1 through  $k + 1$  [7]

Recursive interpretation is:

$$shortestPath(i, j, k) = \min(shortestPath(i, j, k - 1), shortestPath(i, k, k - 1) + shortestPath(k, j, k - 1));$$

$$shortestPath(i, j, 0) = edgeCost(i, j);$$

(ALG 1)[7]

Algorithm starts computing with  $shortestPath(i, j, 1)$  for all pairs  $(i, j)$  and then using  $shortestPath(i, j, 2)$  for all pairs  $(i, j)$  etc. Process continues till  $k = n$ .

Time complexity:

To find all  $n^2$  of  $W_k$  from those of  $W_{k-1}$  requires

$2n^2$  operations. Since we begin with  $W_0 = W_R$  and compute the sequence of  $N$  zero-one

matrices  $W_1, W_2, W_3, \dots, W_n = M_R$  than total

number of operations used is  $n * 2n^2 = 2n^3$ . Therefore the complexity of algorithm is  $\Theta(n^3)$  and can be solved by a deterministic machine in polynomial time. [7]

Function  $W_{ij} = edgeCost(i, j)$  determines complexity of passing from one entity to other. On the basis of attributes of connection and attributes from starting vertex and ending vertex we are able to compute complexity of passing.

Controlling parameters - are facilities for analyst, which uses diagram of security to analysis of risk management and can increase weight of significant attributes.

Controlling parameters are:

- Weight of user knowledge  $WUk$
- Weight of power of facilities  $WUp$
- Weight of number of access  $WUna$
- Weight of software security on entity  $WEss$
- Weight of number of incident on entity  $WEni$
- Weight of data sensitivity  $WDSds$
- Weight of number of incident on data sources  $WDSni$
- Weight of level of cryptography  $WTC$
- Weight of security transmitting data  $WDSec$

User of system  $u \in U$ , which attributes are:

- Knowledge  $Uk$
- Power of facilities  $Up$
- Number of access  $Una$

Attributes of intersection edge  $h_{ij} \in (E \cup DS)$

1. Cryptography level  $TC$
2. Security of transmitting data  $DSec$

In case, that end point vertex is type of entity  $v_j \in E$ , input attributes are:

1. Software security  $Ess$
2. Number of incidents  $Eni$

Formula for compute result complexity (weight):

$$W_{ij} = (U_k * WU_k + U_p * WU_p + U_{na} * WU_{na}) * (E_{ss} * WE_{ss} + E_{ni} * WE_{ni} + TC * WTC + DSec * WDSec) \quad (F1)$$

In case, that end point vertex is type of data source  $v_j \in DS$  input attributes are:

1. Data sensitivity  $DSds$
2. Number of incidents  $DSni$

Formula for compute result complexity (weight):

$$W_{ij} = (U_k * WU_k + U_p * WU_p + U_{na} * WU_{na}) * (DS_{ds} * WDS_{ds} + DS_{ni} * WDS_{ni} + TC * WTC + DSec * WDSec) \quad (F2)$$

### B. Display risk in diagram of security

In order to get information about risk is necessary compute reachability tree of entities and data sources from user. Through to graph decomposition to tree structure we can get elements which are accessible for user and evaluate scale of security risk and security of available elements.

I use decomposition algorithm - depth first search algorithm. Depth first search algorithm – that progresses by expanding the first child node of the search tree that appears and thus going deeper and deeper until a goal node is found, or until it hits a node that has no children. Then the search backtracks, returning to the most recent node it hasn't finished exploring. Algorithm write in pseudo programming language is follow:

```
dfs(graph G)
{
  list L = empty
  tree T = empty
  choose a starting vertex x
  search(x)
  while(L is not empty)
  {
    remove edge (v, w) from beginning of L
    if w not yet visited
    {
      add (v, w) to T
      search(w)
    }
  }
}

search(vertex v)
{
  visit v
  for each edge (v, w)
    add edge (v, w) to the beginning of L
}
(Alg2)
```

Time complexity of algorithm is proportional to the number of vertices plus the number of edges in the graphs it traverse  $\Theta(|V| + |E|)$ . [8]

When we apply algorithm in diagram of security by choose starting vertex which is type of users and then make decomposition of diagram of security to tree structure, we can determine paths to entities or to data sources for selected user. When we need determine risk of access to entity or data source we have to extend behaviors of vector of access. For every one user we determine potential risk of access to entity or data source. Computed risk we display on edges of tree structure. We can display scale of security of entities and data sources by compute with formula F1 or F2.

## VI. CONCLUSION

Diagram of security is tool for risk management. Get risk of access to sensitive data by users or identify weak places in model helps managers of organization in their decisions. When join this model with real system for example with protocol which sending actual information from server about attack and users behaviors from security logs can increase precision of computing of risk and security attributes. Diagram of security is potentially tool for digital forensics investigation and can get answers to question who is probably attacker. Diagram of security can be as extension of UML 2.x Deployment diagram [10]. Entity in security diagram has equivalent element named "Node" in deployment diagram. Data source in security diagram has equivalent element named "Artifact" and intersection in diagram security has equivalent in association or dependency in deployment diagram [11]. Diagram of security adds security knowledge of deployed system.

## ACKNOWLEDGMENT

This work was supported by VEGA Grant No. 1/0350/08 Knowledge-Based Software Life Cycle and Architectures.

## REFERENCES

- [1] Vysoký, M: Príspevok k riešeniu bezpečnosti distribuovaných informačných systémov, TU Košice, Máj 2005.
- [2] Stoneburner, G, Goguen, A, Feringa, A Risk Management Guide for Information Technology Systems, <http://csrc.nist.gov/publications/nistpubs/800-30/sp800-30.pdf>, NIST Special Publication 800-30
- [3] Masný, I : Forenzné analýzy – nástroje bezpečnosti [http://www.emm.sk/files/caweek2004\\_emm\\_forenzne\\_analyzy.pps](http://www.emm.sk/files/caweek2004_emm_forenzne_analyzy.pps), CAWeek, 2004
- [4] Corey, V, Peterman, Ch, Sherin, S, Greenberg, M, Booklen, J, Network Forensics Analysis, <http://www.sandstorm.net/support/netintercept/downloads/ni-ieee.pdf>, IEEE INTERNET COMPUTING, December 2002
- [5] IEONG, R FORZA – Digital Forensics Investigation Framework that incorporate legal issues, <http://www.dfrws.org/2006/proceedings/4-Ieong-pres.pdf>, eWalker Consulting Ltd., 2006
- [6] Vidas, T, Cyber-Forensics The Basics, <http://www.certconf.org/presentations/2006/files/WD4.pdf>, CERTConf, 2006
- [7] Floyd–Warshall algorithm, [http://en.wikipedia.org/wiki/Floyd-Warshall\\_algorithm](http://en.wikipedia.org/wiki/Floyd-Warshall_algorithm), Wikipedia, 2009
- [8] Depth-first search, [http://en.wikipedia.org/wiki/Depth-first\\_search](http://en.wikipedia.org/wiki/Depth-first_search), Wikipedia, 2009
- [9] HAVLICE, Zdeněk et al. : Knowledge-based software life cycle and architectures. In: Computer Science and Technology Research Survey. Košice: TU, 2007. s. 47-68. ISBN 978-80-8086-071-4.
- [10] Scott W. Ambler UML 2 Deployment Diagrams, <http://www.agilemodeling.com/style/deploymentDiagram.htm>, 2007
- [11] OMG Unified Modeling Language, Superstructure, V2.1.2, <http://www.omg.org/spec/UML/2.1.2/Superstructure/PDF/>, 2007

# Metalevel Driven Evolution of Software Languages

Lubomír WASSERMANN

Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

lubomir.wassermann@tuke.sk

**Abstract**—Reflective system (a system that reasons about itself) consists of two subsystems - domain system and metalevel system. Domain system is representing application logic. Metasystem describes controls or modifies objects of the domain system. It contains set of operations called Metaobject protocol (MOP). The main goal of designing and developing MOP is to achieve extensibility of the language. That is the reason why MOP is suitable for evolution of language. Change of language can be realized by appropriate MOP operation. Very important objective is to preserve compatibility and consistency between evolution steps. Each evolution step should be properly documented to keep track of all the changes made to the language. Transformation of language is based on its description generated and realized by MOP and its operations.

**Keywords**—metaobject protocols, metaprogramming, language evolution, reflection

## I. INTRODUCTION

Nowadays, software development is in a big crisis. It is challenging due to two main factors – complexity and change [1]. Software products during their lifecycle are coping with change. Change causes software product to evolve. But to maintain quality of software product, evolution should be controlled in some way with help of language mechanisms and constructs. Traditionally, languages have been designed to be viewed as set of black box abstractions [2]. End programmers have only part or no control at all over the semantics or implementation of these abstractions. This point of view is related to misconception that software languages are immutable [12]. Programs are dependent on language, in which they are written, and tools which this language offers (e.g. interpreters, compilers, etc.). When we admit that program is subject to evolution and is tightly coupled with language in which is written we can assume that language can be subject to evolution, too.

Important aspect of language evolution is to preserve compatibility of programs written in different versions of language. To achieve this, each evolution step need to be properly documented to keep track of all the changes made to the language. “Opening up” language abstractions and implementation could be one way how to achieve language evolution. But as stated in [2], this opening should not expose unnecessary details and thus overwhelm programmer. Only the essential structure of the implementation should be exposed. Providing an open implementation can be advantageous in a wide range of high-level languages. Metaobject protocol (MOP) technology is a powerful tool for providing this [2]. The main goal of designing and developing

MOP for language is to achieve its extensibility. Thus well-designed MOP can serve as tool for language evolution.

## II. METAPROGRAMMING AND METASYSTEMS

As the prefix “meta” is suggesting (meta = being about), metaprogramming is writing programs that represent and manipulate other programs (or programs that write programs). The most common metaprogramming tool is a compiler [13]. To define metasystem we can come out from definition of computational system. According to [3] computational system is a system that acts and reasons about a domain (Fig. 1a). With this definition we can define metasystem as a system whose domain is another computational system (Fig.1b). The computational system which acts as domain of a metasystem is called its base system [4]. Due to fact that every change in domain is reflected in its computational system (causal connection) we can see that every change in computational system is reflected in metasystem reasoning about that computational system, and vice versa. The computational system operates at domain level. It contains application logic represented by domain objects. Metasystem operates at meta-level. It contains system metaobjects, which describe, control or modify domain objects [5]. Metaobject is an object which reflects the structural, and possibly also the behavioral aspect of a single object. Meta-level provides information about selected system and makes the software self-aware.

We can easily introduce reflection to all these definitions when we regard the fact that metasystem is computational system, too. Then computational system which reasons and acts about itself is called reflective system (Fig.1c). Reflective program is a program describing a system that accesses its own metasystem [4].

### A. Reflection

Reflection is an entity's integral ability to represent, operate on, and otherwise deal with itself in the same way that it represents, operates on, and deal with its primary subject matter.

Reflection is used as one of tools of metaprogramming. It is important aspect of relationship between domain object and its metaobject. This relationship allows domain object to ask for the services of metaobject and metaobject to change domain object implementation. Meta-level control over domain level takes part in two steps (Fig. 1) [6], [7]:

- Domain object calls metaobject requesting change in term of semantics. This is called reification. Reification

is process of making concrete an implicit aspect of an object, that it can be changed by the metaobject.

- Flow of control is returned back to domain object. Because the the metaobject modified a part of domain object, its behavior and/or structure is now changed. This process is called reflection – reflecting the changes back to domain object.

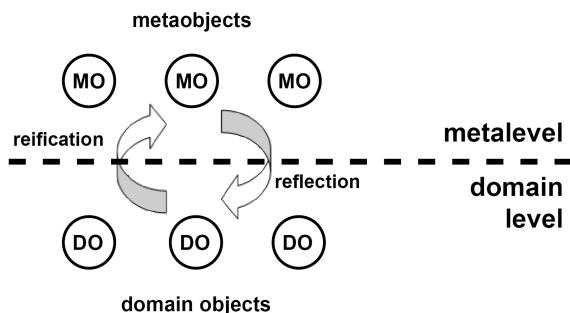


Fig. 1. Processes of reflection

As defined in [8], a reflective mechanism is any means or tool made available to a program P written in a language L that either reifies the code of P or some aspect of L, or allows P to perform some reflective computation. When system reifies some parts of program, there is variety of possible actions over these reifications. Here comes to difference between reflective mechanisms – introspection and intercession.

Introspection is the ability of a program to simply reason about reifications of otherwise implicit aspects of itself or of the programming language implementation (processor). In analogy with file systems, introspection can be seen as a read access to reifications.

Intercession is the ability of a program to actually act upon reifications of otherwise implicit aspects of itself or of the programming language implementation (processor). Following the same analogy, intercession corresponds to a write access to reifications [4].

### B. Metaobject protocols

Metaobject protocol (MOP) can be described as an interface through which domain objects and metaobjects are communicating. This interface can be seen as standard interface between objects but transposed in the area of reflection and metaprogramming. This interface allows independent development of domain system and metasystem.

According to definition stated in [10], Metaobject protocols are interfaces to the language that give users the ability to incrementally modify the language’s behavior and mplementation, as well as the ability to write programs within the language. This definition was later refined and describes three principles of how MOPs work:

1. The basic elements of the programming language (classes, methods and generic functions) are made accessible as objects. They are given the special name of metaobjects because these objects represent fragments of a program.
2. Individual decisions about the behavior of the language are encoded in a protocol operating on these metaobjects.

3. For each kind of metaobjects, a default class is created, which lays down the behavior of the default language in the form of methods in the protocol.

### C. Methodology of MOP design

When we are designing MOP, we can look at it as we were designing a programming language. (comparison is depicted in Fig. 2). Language designers are working with two different levels of design process at the same time - the level of designing particular programs in terms of a given language, and the level of designing the language to support the lower-level design processes.

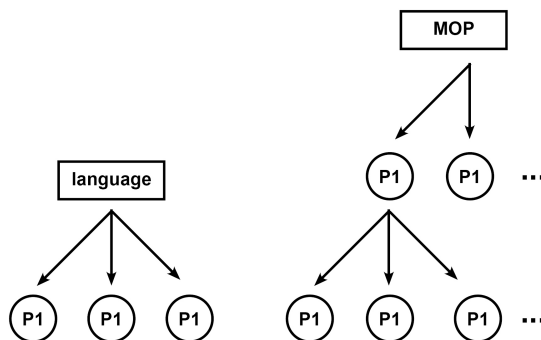


Fig. 2. Language design vs. MOP design

MOP design is similar, with the addition of one more level in design process. Again, designer is considering programs that could be written in language but is not thinking about only one language but a whole range of languages that can be expressed with designed MOP.

## III. EVOLVING LANGUAGE

Well designed language MOP can be used to modify the language and its behaviour. We can assume that MOP is suitable tool to support the evolution of language. Particular MOP operations can change different aspects of language (Fig. 3). But as mentioned before, to avoid inconsistency, evolution of language should be controlled. Imagine that company has developed systems that use domain specific language (DSL) designed by them. New requirements on system can cause the need to cover more concepts from given domain. This necessary leads to extension of system to cover new area of given domain. To be able to express new concepts in DSL language, it is possible that language would have to be extended by new constructs and/or concepts, or altering existing ones. But due to fact that DSL language is used not only by one system, change in language can affect functionality of other systems. To keep the consistency, each evolution step should be documented and transformation of language described to be able to map previous version of language to new version of language (e.g. mapping constructs, elements, keywords, etc.).



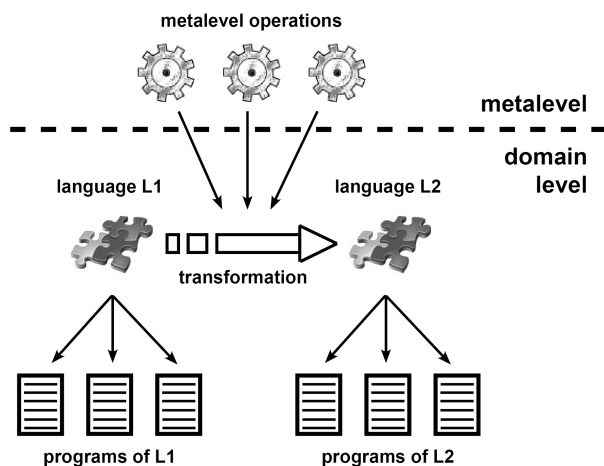


Fig. 3. Scheme of evolution of language

If there is a need for change of language appropriate MOP operation can be called. The result will be language transformation and its description. This description can be used also as documentation of versioning of system or language itself. Whole transformation can be performed at metalevel. Change of language can be achieved by reflective mechanisms of introspection and intercession.

Imagine we have our own DSL language used to write programs which serve as input for our system from medical domain. Programs written in DSL language are input for language processor (compiler or interpreter) which processes them. At metalevel is metasystem with our designed MOP protocol with set of operations that can be used to transform our DSL language. Domain system and metasystem are written in Java. If we want to transform our language that existing concept *Pill* will be renamed to *Medicine* and will contain new property *Type*, appropriate MOP operations are called and description of transformation is generated (Fig. 4). Transformation of language will be based on its description. Metasystem then alter language processor to accept new or modified concepts.

```
<transform>
  <source>
    <concept>
      <name>Pill</name>
    </concept>
  </source>
  <target>
    <concept>
      <name>Medicine</name>
      <property action="add">
        <name>Type</name>
      </property>
    </concept>
  </target>
```

Fig. 4. Description of transformation in XML

Alteration of language processor itself can be done with help of some Java reflection extension or framework. For example, reflective extension JavAssist enables reification and alteration of existing classes and creation of new classes. JavAssist and its functionality is based on bytecode transformation [9].

Language processor is transformed in two ways. First, as mentioned before, language processor have to conform to new

language syntax, to be able to process programs in new version of language, processing new concepts and constructs or altered concepts.

Second, language processor should be able to process programs written in previous version of language. This program is transformed and processed according to description of transformation of language. The reason why we stressed the necessity of description of transformations and documenting them as part of system (or language) versioning is the situation when we need to know history of language evolution steps and their transformation. Program is then processed and transformed according to history of evolution steps.

This way the language processor would be able to process not only programs of recent version of language. Integrity and functionality of systems that share same DSL language which is evolving through time will not be compromised. Controlled evolution of language will ensure us better coping with change of software systems.

#### IV. CONCLUSION

In this paper we have presented approach to language evolution with use of metaobject protocol. We were concentrating on preservation of consistency and compatibility between programs written in different “versions” of evolving language. Each evolution step and its language transformation is properly described. This description of transformation is input for transformation of language processor to accept new or modified concepts. Description of transformation is also used for mapping between concepts of previous version of language with concepts of new version.

#### ACKNOWLEDGMENT

This work was supported by VEGA Grant No. 1/4073/07 – Aspect-oriented Evolution of Complex Software System.

#### REFERENCES

- [1] E.Althammer: Reflection patterns in the context of object and component technology, University of Konstanz, PhD thesis, 2002, 120 pp.
- [2] M. Bebjak, V. Vranić, and P. Dolog. Evolution of Web Applications with Aspect-Oriented Design Patterns. In Marco Brambilla and Emilia Mendes, editors, Proc. of ICWE 2007 Workshops, 2nd International Workshop on Adaptation and Evolution in Web Systems Engineering, AEWSE 2007, in conjunction with 7th International Conference on Web Engineering, ICWE 2007, July 19, 2007, Como, Italy, pp. 80–86.
- [3] D. da Cruz, M. Berón, P.R. Henriques, and M.J.V. Pereira: Strategies for Program Inspection and Visualization, Proceedings of CSE 2008, International Scientific Conference on Computer Science and Engineering, Sep.24–26, 2008, High Tatras, Slovakia, pp. 107–117.
- [4] J.M. Favre: Languages evolve too! Changing the Software Time Scale, Proceedings of the Eighth International Workshop on Principles of Software Evolution, 2005, pp. 33–44.
- [5] D. Friedman and M. Wand: Reification: Reflection without metaphysics, In Proceedings of the Annual ACM Symposium on Lisp and Functional Programming, August 1984, pp. 348–355.
- [6] J. Greenfield, K. Short, S. Cook, and S. Kent: Software Factories: Assembling Applications with Patterns, Models, Frameworks, and Tools, Wiley, 2004, 500 pp.
- [7] G. Kiczales, J. Rivieres, and D. Bobrow: The Art of the Metaobject Protocol, MIT Press, 1991, 236 pp.

- [8] G. Kiczales et al.: *Metaobject Protocols: Why We Want Them, and What Else They Can Do In Object-Oriented Programming*, Andreas Paepcke, Ed., MIT Press, Cambridge, MA, 1993, pp. 101–118.
- [9] J. Kollár, J. Porubán, P. Václavík, J. Bandáková, and M. Forgáč: *Software Evolution From A Meta-Level Compiler Perspective*, SCIENCE & MILITARY, 2, 2, 2007, pp. 29–32.
- [10] P. Maes: *Computational reflection*, Ph.D. thesis, Artificial intelligence laboratory, Vrije Universiteit, Brussels, Belgium, 1987., 112 pp.
- [11] J. Malenfant, M. Jacques, and F. Demers: *A tutorial on behavioral reflection and its implementation*, In *Metaobject Protocols*, Kiczales, Ed., MIT Press, Cambridge, MA, 1996, pp. 1–20.
- [12] R. Pawlak: *Metaobject Protocols For Distributed Programming*, Technical report, Laboratoire CNAM-CEDRIC, Paris, 1998, 15 pp.
- [13] J. Porubán and P. Václavík: *Generating Software Language Parser from Domain Classes*, Proceedings of CSE 2008 International Scientific Conference on Computer Science and Engineering, The High Tatras – Stará Lesná, Slovakia, Sep. 24–26, 2008, pp. 133–140.
- [14] E. Tanter: *From Metaobject Protocols to Versatile Kernels for Aspect-Oriented Programming*, PhD thesis, University of Nantes, France, and University of Chile, Chile. November 2004, 104 pp.
- [15] P. Václavík and J. Porubán: *Template-based content management system*, AEI'2008 International Conference on Applied Electrical Engineering and Informatics, Athens, Greece, September 8–11, 2008, pp. 153–157.
- [16] C. Zimmermann: *Advances in Object-Oriented Metalevel Architectures and Reflection*, CRC Press, 1996, 336 pp.
- [17] D. Zmaranda, G. Gabor: *Software Environment for Task oriented Design of Real Time Systems*, Proceedings of CSE 2008, International Scientific Conference on Computer Science and Engineering, Sep.24–26,2008, High-Tatras, Slovakia, pp. 359-366.
- [18] D. Zmaranda, G. Gabor, M. Gligor: *A Framework for Modeling and Evaluating Timing Behaviour for Real-Time Systems*. Proc. SINTES 12 Int. Symposium on Systems Theory, Oct. 20–22, University of Craiova, Romania, 2005, pp. 514-520.

# Information Systems Architectures Driven by Project-knowledge

*Peter Žárský*

Dept. of Computer Science and Informatics, FEI TU of Košice, Slovak Republic

peter.zarsky@tuke.sk

**Abstract**— This paper is about the concept of an architecture, which integrates project knowledge as a part of architecture of information system.

Project knowledge have important role for the management, maintenance and the modification of complex software systems. There are many circumstances during life cycle of information system, when from many reasons, documentation became inconsistent with the real status of the information system. The proposed approach aims to eliminate this problem and to facilitate the management and the maintenance of information systems. Besides the integration of the project knowledge directly into the information system's architecture of in a form that is extractable by user or external tool form management of information system, the proposed concept of this architecture offers additional advantages. This architecture allows upgrading or replacing parts of the information system without the necessity of breaking the system's operation. It enables self-adaptation of the systems' structure. In addition, all changes in the information system will take affect in documentation and vica-versa, all changes in model of the information system will affect the behavior of the information system.

**Keywords**— architecture, documentation, information system, project knowledge

## I. MOTIVATION

During the development and the maintenance of information systems it happens that there are differences between the models and descriptions defined in the documentation and the actual phase of its implementation. It is caused by separated evolution of the information system and its documentation. This leads to the discrepancy in the usage of the information system. Inconsistence between the documentation and implementation causes that information system appears to be a white-box only to its developer. Dependency on the responsible developer is becoming dangerous, when she/he leaves the company. This fact motivated, to create an architecture that will integrate project knowledge directly into the architecture of the information system itself in an indivisible way. In addition, it has to enable reading and maintaining this knowledge by a user or external tool that could help in reengineering the model of the information system.

## II. BASES OF THE PROPOSED ARCHITECTURE

We approach this concept using the following three approaches: Component-Based Architecture (CBA), Service-Oriented architecture (SOA), Model Driven Architecture (MDA).

### A. Component-Based Architecture

Application based on CBA is formed from components. Component is the executable part of source code . CBA consist from specification, source code and executable code. Each component must have the specified interface, that enables communicate with other components to form more complex entities. Motivation in using CBA is in the possibility of components reusability [1], [2].

### B. Service-Oriented Architecture

In this type of architecture application consists of objects named services accessible on network. In compared with the traditional static software architecture, the architecture of SOA based application is dynamic. It means, that application can be composed in time of program execution. This enables dynamically change the application, to provide maintenance or implement new requirements [3], [4].

### C. Model-Driven Architecture

Model-Driven Architecture is architecture based on Object Management Group's (OMG) established standards. MDA separates business and application logic from the underlying platform technology. MDA based development starts with creating Platform-Independent Model (PIM) of an application's business functionality and behavior. The PIM remains stable as technology evolves. MDA unifies the development of an application from its start as a PIM through one or more Platform-Specific Models (PSMs), to generated code and a deployable application [5].

## III. STRUCTURE OF THE PROPOSED CONCEPT

Structure of this concept consists of three main parts. Set of components that is the core of the information system , base of project knowledge and GUI (Graphical User Interface) based tool. This provides access to the base of the project knowledge

for the designer or user, as shown on Fig. 1.

*A. Set of components of the information system*

This part of information system is the logical set of all components. It includes also external services or resources, that are used by the information system. This set does not consist only of components that are executable, reusable part of code as in CBA. Parts of this set are all data resources used by the information system, configuration files, external services etc. Each component has its specification stored in base of project knowledge. Component itself does not contain any information about an other component that is active or passive communication. This information is read from base of

new components, checking integrity of information system). For this knowledge access time is not critical and for storing them can be used any DBMS (database management system) or even a set of XML files. In the second group, there are the compact knowledge about the relationships and communication channels between all components from the set of components. Compared to the first group of knowledge, this group includes knowledge needed during execution of components and access time is therefore critical. The amount of data stored in this part of information system is not enormous, so they can be cached in main memory for faster access.

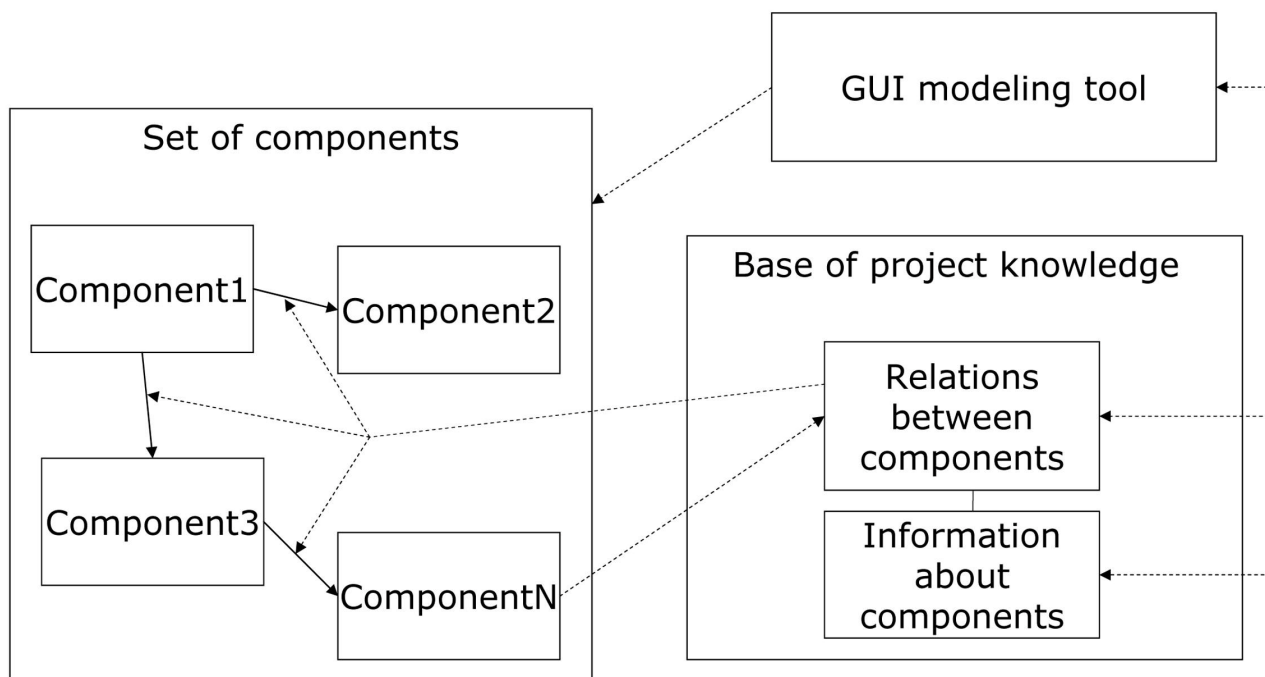


Fig. 1. Structure of architecture described in concept consisting from three parts: set of components, modeling tool, and base of project knowledge.

project knowledge every time it is needed. It is read from the knowledge base at start of the execution of the component or immediately before the knowledge is needed, that depends on the component's demands. This ensures that components behave according to actual model stored in base of project knowledge.

*B. Base of project knowledge*

Key part of the concept for interconnection of implementation and documentation of the information system is base of project knowledge. This part of information system stores knowledge that are used by both: components and GUI based tools that enable to reengineer the model of that information system. Project knowledge is stored in base and can be divided into groups. In the first group, there are stored specifications and notes of all components from set of components. This knowledge are used only for accessing documentation to reengineer component model of information system and during modification of information system (adding

*C. Modeling tool for access to the base of project knowledge through GUI*

The third part of this concept of architecture is tool with GUI that allows to developer or user to view and change model of information system. This tool will reengineer model of information system from base of project knowledge, so it will be always correspond with the current version of the information system and this is the main advantage compared to the standard documentation. This tool should allow adding and removing of components from set of components, maintain relations between components and check integrity of the system. Integrity of the system can be corrupted in two ways. User can try to remove a component that has defined relation with another component. This can lead to situation, when component needed by another component is missing in the system. Another case of corrupting the system integrity is, if user will try to add new component into set of components without defining all relations of this component described in its interface specification. It would cause the same situation as previous case.

## ACKNOWLEDGMENT

This work was supported by VEGA Grant No. 1/0350/08 Knowledge-Based Software Life Cycle and Architectures.

## IV. MAINTENANCE AND MODIFICATIONS

Information system during its life cycle evolves. Requirements to functionality of information system are changing due to changes in the business processes and other circumstances. Advantage of this concept are that for the most changes it is not necessary to stop operation of the information system.

Modifications of the information system based on this architecture can be divided into to groups. Modifications, that relate to the development of new components, and modifications, which can be done on the model stored in the base of project knowledge using existing components.

In the first case, it is necessary to develop one or more components firstly. We have to choose appropriate granularity regarding to reusability of components. Prepared components can be added into information system using modeling tool. In the first step we add new components into the set of components, with its definition. In the second step we modify model of information system and define new relations between the components. Modeling tool will check integrity of the system and in case that integrity is not corrupted, it will make changes in base of the project knowledge.

Those modifications that do not need development of new components can be done by changing model by GUI tool. But this type of modification can be done also by information system itself on command of user or as reaction on an event etc. Very simple example is switching on or switching off logging of some event by adding or removing of component, which is executing this task. Advantage of this approach, compared to changing some variable in configuration file etc., is, that this change will be visible also in model in modeling tool. All changes, made by information system, must be programmed carefully to avoid corruption of integrity of information system.

Another advantage of this concept is that it allows modification of system without stopping operation of system. Even if we want to replace binary code component with new version, there is no need to stop executing of this component. New version of component can be added into system as a new component and both versions of component will coexist as two independent components. Model will be modified to redirect all new queries to new version of component, but old version of component will continue its operation until it finishes all tasks. When the old version of component will stop, it can be removed from the system.

## V. CONCLUSION

Integration of project knowledge into architecture change character of information system from black-box into white-box. Possibility to access project knowledge without the need of recompilation brings other advantages in maintenance and further modifications. Despite all advantages, it does not mean that this concept is replacing standard documentation, it is only complementing it.

## REFERENCES

- [1] Component-based Architecture And Modeling and Simulation, Keane, NDIA, In Proc. of 5<sup>th</sup> Simulation-Based Acquisition/Advanced Systems Engineering Environment Conf., June 24-27, 2002
- [2] Lüders F.: Use of Component-Based Software Architectures in Industrial Control Systems, Mälardalen University Licentiate Thesis No. 18, Department of Computer Science and Engineering, Mälardalen University, (c) Frank Lüders, 2003, ISBN: 91-88834-19-0, Printed by Arkitektkopia, Västerås, Sweden, Distribution: Mälardalen University Press.
- [3] Sprott D., Wilkes L.: Understanding Service-Oriented Architecture, CBDI Forum, The Architecture Journal, JOURNAL, 1, accessible 8.3.2009 on [http://download.microsoft.com/download/4/d/a/4da0ddee-77e0-47d1-aaa7-a5dd619b8bca/journal1\\_english.pdf.zip](http://download.microsoft.com/download/4/d/a/4da0ddee-77e0-47d1-aaa7-a5dd619b8bca/journal1_english.pdf.zip)
- [4] Beličák M.: Inteligentná komponentová architektúra informačných systémov, dissertation work
- [5] Object Management Group, Model Driven Architecture Specifications accessible 2.3.2009 on <http://www.omg.org/mda/specs.htm>

## Author's index

### A

Adamuščinová Iveta 111  
Aftanas Michal 14  
Andoga Rudolf 245

### B

Baník František 114  
Bánoci Vladimír 18  
Barto Michal 245  
Béreš Tomáš 22  
Blichá Radovan 25  
Bodor Marcel 87  
Bratrů Peter 117  
Bugár Gabriel 18

### C

Cabúk Pavol 103  
Cibuľa Ľubomír 27

### Č

Čačková Viera 31, 34, 53

### D

Dedinská Lýdia 34, 31, 53  
Drotár Peter 119

### Ď

Ďurčík Zoltán 122  
Ďurišin Juraj 38

### E

Eperješi Juraj 126

### F

Fecilak Peter 154  
Fedor Zlatko 130  
Fífik Martin 41  
Forgáč Michal 134

### G

Gazda Juraj 137

### H

Hládek Daniel 141  
Hlubeň Daniel 43

### Ch

Chovaňák Juraj 145

### J

Jancurová Lucia 148, 151  
Janitor Jozef 154  
Jeleň Vladimír 157  
Jenčík Marián 160

### K

Karch Peter 162  
Kažimír Ján 166  
Klimek Ivan 168  
Kliment Ján 173  
Kocan Pavol 176, 64  
Kolesárová Anna 47  
Kopčo Norbert 245  
Kováč Michal 49  
Köver Štefan 178  
Kunštár Ján 182  
Kuzma Miron 186  
Kvakovský Milan 53, 31, 34

### L

Lakatoš Matej 190  
Lapko Marek 193  
Lizák František 57  
Lojka Martin 196, 72  
Lonščák Richard 200  
Lukáč Gabriel 204

### M

Madoš Branislav 207  
Medved' Dušan 60  
Michalko Miroslav 210  
Modrovičová Jana 213  
Mochňáč Ján 64, 176  
Molnár Ján 66

### N

N.Kovács Attila 217, 259  
Nasr Maher 69

### O

Olejár Martin 22

### P

Papaj Ján 221  
Papco Marek 72, 196  
Perhác Ján 224  
Peťko Ivan 227  
Poór Peter 75  
Popovič Ľuboš 79

### R

Reiff Tomas 130  
Ročkai Viliam 231  
Rovňáková Jana 83  
Ruščin Vladimír 87

### S

Sabo Miroslav 234  
Sakmár Michal 90  
Semančík Peter 92  
Staš Ján 238

### Š

Šesták Kristián 242  
Šterba Ján 95  
Švecová Mária 99

### T

Tokár Tamás 246  
Tomoriová Beáta 249  
Tutoky Gabriel 252

### V

Val'a Martin 151  
Vehec Igor 103  
Verbová Anita 256  
Vince Tibor 107  
Vrana Jozef 260  
Výrost Marek 263, 217  
Vysoký Marek 267

### W

Wassermann Ľubomír 271

### Ž

Žárský Peter 275

**9<sup>th</sup> Scientific Conference of Young Researchers  
of Faculty of Electrical Engineering and Informatics  
Technical University of Košice**

Proceedings from Conference

Published: Faculty of Electrical Engineering and Informatics  
Technical University of Košice

I. Edition, 279 pages, the number of CD Proceedings: 120 pieces

Editors: prof. Ing. Alena Pietriková, PhD.  
Ing. Attila N.Kovács  
Ing. Jana Modrovičová

**ISBN 978-80-553-0178-5**