Scientific Conference of Young Researchers 2013

13th Scientific Conference of Young Researchers

May 14th, 2013
Herľany, Slovakia

Proceedings from Conference

Faculty of Electrical Engineering and Informatics
Technical University of Košice




Sponsors


TECHNICAL UNIVERSITY
OF KOŠICE
FEI


SES


elfa
nové myšlienky
nové možnosti


ZVT

13th 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, 448 pages, 2013
number of CD Proceedings: 150 pieces

Editors: Prof. Ing. Alena Pietriková, CSc.
Ing. Milan Nosáľ
Ing. Ivan Halupka
Ing. Emília Pietriková

ISBN 978-80-553-1422-8

Program Committee of 13th Scientific Conference of Young Researchers of Faculty of Electrical Engineering and Informatics Technical University of Košice

General chair:	Prof. Ing. Liberios Vokorokos, PhD.
Chairman:	Prof. Ing. Alena Pietriková, CSc.
Members:	Prof. Ing. Roman Cimbala, PhD. Assoc. Prof. Ing. Ján Gamec, CSc. Assoc. Prof. Ing. Želmíra Ferková, CSc. Prof. Ing. Jozef Juhár, CSc. Prof. Ing. Ján Paralič, PhD. Assoc. Prof. Ing. Jaroslav Porubän, PhD. Assoc. Prof. Ing. Ladislav Samuelis, CSc.

Organization Committee of 13th Scientific Conference of Young Researchers of Faculty of Electrical Engineering and Informatics Technical University of Košice

Members:	RNDr. Peter Duranka Ing. Tibor Rovenský Ing. Martina Tarhaničová Ing. Radoslav Bielek Ing. Jozef Lipták Ing. Martin Bačko Ing. Milan Nosáľ Ing. Ivan Halupka Ing. Emília Pietriková Ing. Vieroslava Sklenárová Ing. Radovan Sivý Ing. Martin Broda Mgr. Jana Petrillová Ing. Kornel Ruman
Secretary:	Ing. Ivana Olšiaková
Contact address:	Faculty of Electrical Engineering and Informatics Technical University of Košice Letná 9 042 00 Košice Slovak Republic

Foreword

Dear Colleagues,

SCYR (Scientific Conference of Young Researchers) is a Scientific Event focused on exchange of information among young scientists from Faculty of Electrical Engineering and Informatics Technical University of Košice – series of annual events that was founded in 2000. Since 2000 the conference has been hosted by FEI TUKE with rising technical level and unique multicultural atmosphere. The 13th Scientific Conference of Young Researchers (SCYR 2013), conference of Graduates and Young researchers, was held on 14th May 2013 in Herľany. The primary aims of the conference, to provide a forum for dissemination of information and scientific results relating to research and development activities at the Faculty of Electrical Engineering and Informatics has been achieved. 102 participants mostly by doctoral categories were active in the conference.

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, electrotechnics, computers and informatics, cybernetics and Artificial intelligence, electric power engineering, remained unchanged.

13th Scientific Conference of Young Researchers at Faculty of Electrical Engineering and Informatics Technical University of Košice (SCYR 2013) was organized in the Education and Training Establishment in Herľany of Technical University of Košice. The Conference was opened in the name of dean prof. Ing. Liberios Vokorokos, PhD. by the vicedean of faculty, prof. 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 traditionally includes two parallel sessions (both consist of oral and poster part):

- Electrical & Electronics Engineering (EEE): experimental and theoretical work in the wide-spread field of electronics and electrical engineering,
- Informatics & Telecommunications (IT): experimental and theoretical work in the wide-spread field of informatics and telecommunication,

with about 117 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 fulfilment 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 among 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.

Prof. Ing. Liberios VOKOROKOS, PhD.

General chair of conference and Dean of FEI

Prof. Ing. Alena Pietriková, CSc.

Chair of conference

2013 May 15th, Košice

Contents

Informatics & Telecommunications – Oral form

Informatics & Telecommunications – Poster form

Electrical & Electronics Engineering – Oral form

Electrical & Electronics Engineering – Poster form

Section:
Informatics & Telecommunications
Oral form

Audio Stream Discrimination via Rule-based and Model-based Approach

Jozef VAVREK (3^{rd} year)
Supervisor: Anton ČIŽMÁR

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

jozef.vavrek@tuke.sk

***Abstract*—The comparison between two general classification architectures is investigated in this paper. The first one, also known as the rule-based architecture, utilizes the set of adopted features in order to discriminate between speech/non-speech and music/environment sound. A simple technique based on the threshold setting is implemented. Much more complex architecture utilizes the model-based approach in which a discrimination function is modeled by a decision function or probabilistic models, generated by a classifier. The one-against-one support vector machines (OAO-SVM) and the binary decision trees (BDT-SVM) classification schemes are evaluated in proposed architecture. The experimental results show that the combination of rule-based approach and the OAO-SVM classification scheme reduces the overall miss-classification rate that propagates from the block of speech/non-speech discrimination. Therefore, it performs better than the BDT-SVM architecture.**

***Keywords*—Audio stream classification, rule-based classification, model-based classification, support vector machines.**

I. Introduction

One of the crucial process in audio classification is to discriminate between each acoustic event within the audio content. The main focus is concentrated on speech/non-speech discrimination. Therefore, there is a need to search through frequency and time domain of the signal amplitude and find parameters that capture all the spectral and temporal variations, which represent the audio content of speech and non-speech sounds. The optimal parameterization of audio data is also very important for applications such as broadcast news (BN) transcription and retrieval, where speech signals are typically mixed with music and environment sound alternately. Processes mentioned above are inevitable for building an appropriate audio segmentation and classification architecture. This paper is therefore focused on content-based classification of BN audio data utilizing rule-based principles and SVM classification architecture.

Audio stream of BN data comprises five general acoustic classes, namely *pure speech*, *speech with environment sound*, *environment sound*, *speech with music* and *music*. Each individual class is characterized by acoustic properties related to speech and non-speech audio events and random occurrence in audio stream. BN audio stream also contains silent intervals between different speakers and jingles, defined as *silence*.

A comprehensive research has been carried out in the field of audio classification by employing different features and techniques. Most of them are aimed on the design of such complex systems that are capable to classify the audio stream into speech and non-speech segments by using novel features and rule-based classification approach. For example, Wang et al. [1] proposed simple but robust method to discriminate speech and music by implementing the modified version of low energy ratio, the Bayes MAP classifier and a novel context-based post-processing technique based on the four state transition model. The classification accuracy of more than 97% has been achieved despite the low computational complexity of their method.

Lu et al. [2] proposed unique two-stage audio segmentation and classification algorithm capable of segmenting and classifying an audio stream into speech, music, environment sound and silence by using a set of new features such as noise frame ratio and band periodicity. Experimental evaluation has shown high classification efficiency, over 96%, even with a smaller testing unit.

Kos et al. [3] used a novel feature for online speech/music segmentation based on the variance mean of filter bank energy (VMFBE). Different experiments, including evaluations on different audio databases (BNSI and radio broadcast database) and comparison with other well known audio features such as spectral centroid (SC), spectral roll-off (SR), percentage of low energy frames (PLEF), SF, ZCR and mel-frequency cepstral coefficients (MFCC), have been performed with a reported overall classification accuracy of over 94% within $200\ ms$ segments.

Much more competitive task is to discriminate voiced and unvoiced speech from its background interference. Han and Wang [4] proposed a classification approach utilizing SVM and pitch-based features in combination with amplitude modulation spectrum features. The performance of proposed system was compared with Kim et al. system [5] and experimental results lead to significantly improved speech intelligibility in noisy conditions.

This paper is organized as follows: Section II deals with content-based analysis and feature extraction methods. Section III provides description about used classification architecture and Section IV discusses realized experiments and obtained results. Finally, the Section V gives our conclusions and shows future directions.

II. Content-based analysis

The content-based analysis helps to extract desired parameters from the audio signal waveform. These parameters, also called *features* or *descriptors*, capture temporal and spectral characteristics of the audio signal and characteristics of

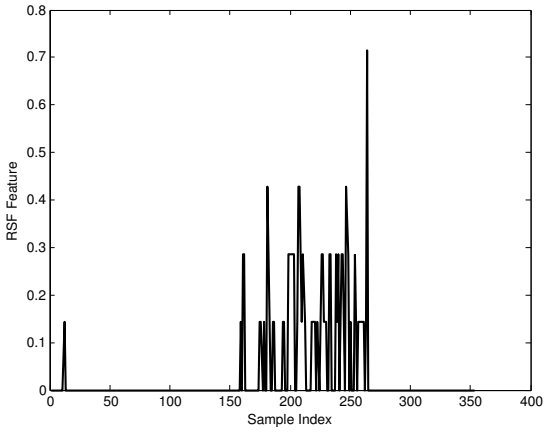


Fig. 1. Ratio of silent frames curve; 0-160 samples music, 161-270 samples speech, 271-350 samples environment sound

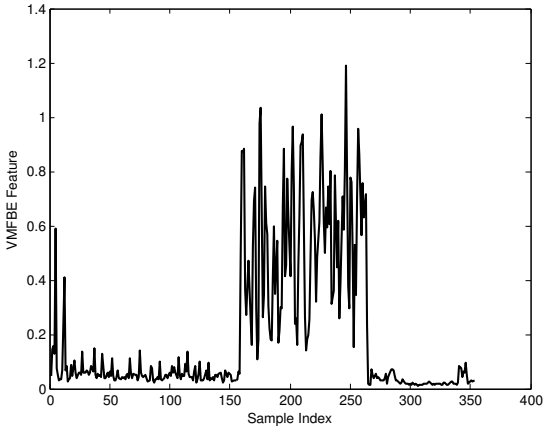


Fig. 2. Variance mean of filter bank energy curve; 0-160 samples music, 161-270 samples speech, 271-350 samples environment sound

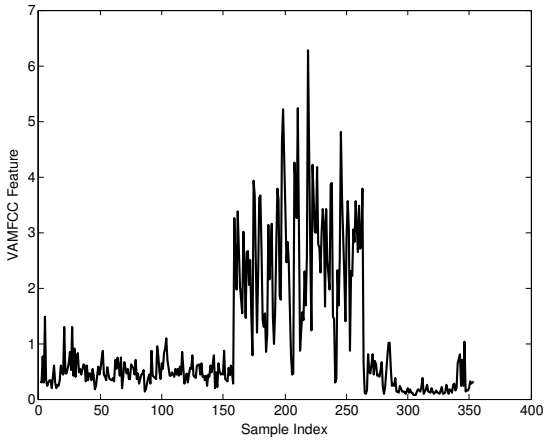


Fig. 3. Variance of acceleration MFCCs curve; 0-160 samples music, 161-270 samples speech, 271-350 samples environment sound

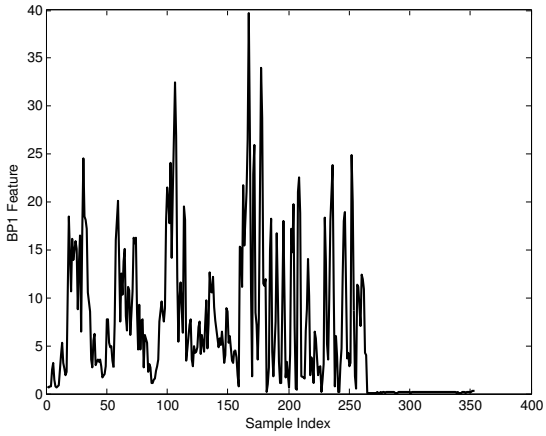


Fig. 4. Band periodicity curve; 0-160 samples music, 161-270 samples speech, 271-350 samples environment sound

the speaker vocal tract. Therefore, the following subsections describe the audio descriptors, we decided to use in our classification architecture, due to their strong discrimination ability between speech and non-speech sounds.

A. Ratio of Silent Frames

Ratio of Silent frames (RSF) [6] measures the proportion of silent frames exploiting the energy contour of an audio signal waveform and the number of times the audio waveform crosses the zero axis (ZCR). RSF appears to be an effective descriptor, especially for speech/non-speech discrimination. For non-speech signals, the contour tends to show minimal change over a period of several seconds, but there are considerable changes between voicing and short pauses in speech signals. It follows the fact, that there are more silent frames in speech than in music segments, what can be seen clearly from the characteristic depicted on Fig. 1.

B. Variance Mean of Filter Bank Energy

VMFBE feature [3] calculates an energy variation in a narrow frequency sub-band of signal's spectrum. It exploits the fact that energy varies more rapidly and to a greater extent for speech than for non-speech. Therefore, an energy variance in such a sub-band is greater for speech than for music or environment sound. The process of computing the VMFBE features consists of the following steps:

1. Divide the signal into the overlapping frames, using Hamming window function, and compute the fast Fourier transform (FFT).
2. Calculate the logarithmic energy in each sub-band of mel-filter bank by multiplying each FFT magnitude coefficient with the corresponding sub-band filter channel gain.
3. Compute the variance of an individual filter channel:
4. Calculate the mean of these coefficients:

Representation of the VMFBE feature, exploiting 24 filter bands, can be seen on Fig. 2.

C. Variance of Acceleration MFCCs

Mel frequency cepstral coefficients (MFCCs) are considered as the most effective features, especially in the task of speech recognition and speech/non-speech discrimination [7]. The calculation procedure of the MFCC features is almost similar to the VMFBE feature computation except for the sub-band filter energy variance and the energy variance mean calculation. This step is replaced by computing the Discrete Cosine Transform (DCT). Usually, only the first 12 cepstrum coefficients (excluding the 0th energy coefficient) are used. But they do not include a temporal information. The variability of speech signals is much more significant in comparison with music signals. Therefore we decided to use variance of acceleration MFC coefficients (VAMFCC) to assess the temporal dynamics of audio signals and as a robust speech/non-speech discriminator (Fig. 3).

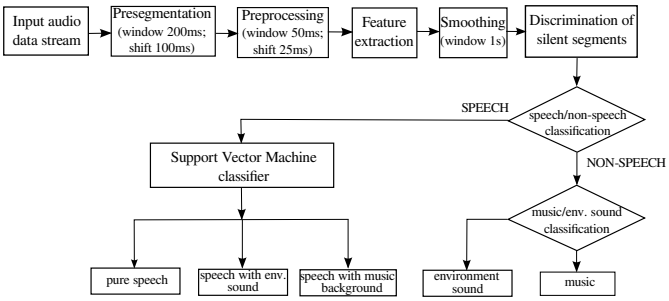


Fig. 5. Classification scheme of the speech-music-environment sound discrimination system

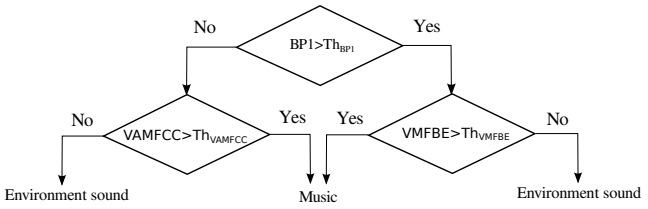


Fig. 6. Music/environment sound classification process

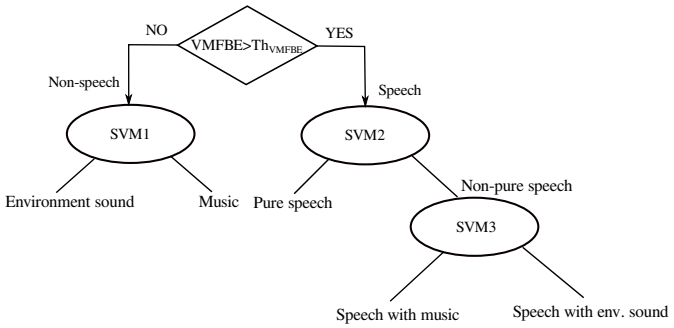


Fig. 7. Binary decision trees classification architecture utilizing SVM

D. Band Periodicity

Band periodicity (BP) [2] is defined as the periodicity of particular frequency sub-band and can be represented by the maximum local peak of the normalized correlation function within each sub-band. The bandwidth selection of each sub-band is conditioned by the sampling frequency and the frequency range of the audio signal we want to examine. The most energy of frequency components that represent speech signal is concentrated at lower frequency bands, while some frequency components of music signals are much more significant at higher sub-band frequencies. Therefore, we decided to examine the following four sub-bands $500 - 1000\ Hz$, $1000 - 2000\ Hz$, $2000 - 4000\ Hz$ and $4000 - 8000\ Hz$, utilizing 20th-order Chebyshev bandpass filter. Fig. 4 shows an example of band periodicity for the first sub-band. There can be observed relatively strong discrimination power between music and environment sound.

III. Classification Architecture

Used classification architecture is depicted on Fig. 5. The block of *pre-segmentation* divides a continuous audio stream into short audio portions with an equal length, also known as *segments*, by using rectangular window. The length of each segment was set to 200 ms with 100 ms overlapping. Each segment is further divided into the overlapped frames, using Hamming window and pre-emphasized by a FIR filter in block of *preprocessing*. The length of each frame was set to 50 ms with 25 ms overlapping in our implementation. All the features presented in previous section are then calculated within each individual frame in time, frequency or cepstral domain. The output of the *feature extraction phase* is represented by a feature vector matrix. Such representation of audio signal defines the format of input data for a classifier.

The process of *smoothing* can help to alleviate the influence of the abrupt changes between several adjacent coefficients within the feature vector that represents only one audio class and, as a consequence of that, reduces the classification error. It is a simple technique based on the averaging of values for a particular feature set within the 1 s floating window.

There is also a need to *remove silent segments* from the audio content. This step can be very useful in reducing the amount of input data and decreasing the computation complexity and training time in the process of classification. The main disadvantage of using silence discriminator is relatively high miss-classification error caused by incorrectly classified silence and non-silence segments. Therefore, we decided to skip this discrimination step due to low discrimination ability of used features (avg accuracy of VAMFCC 82.38%) and to consider only audio data without silent parts excluded manually.

The primary speech/non-speech and music/environment sound classification task was performed by using *rule-based classification*. Such classification approach utilizes the basic principles of discrimination between two events by using the empirical threshold of the observed feature. A criterion for selecting an optimal feature for each classification task was derived from the smoothing probability distribution of speech, music and env. sound histograms. It helped us to choose the range of possible thresholds within each used feature (VMFBE, VAMFCC, RSF and BP). Consequently, the average accuracy of two classes (speech/non-speech and music/background) was used as evaluation criterion in order to set the final threshold of each feature:

$$Accuracy = \frac{1}{2}\left(\frac{TP}{TP+FP} + \frac{TN}{TN+FN}\right), \qquad (1)$$

where TP and TN represent number of correctly classified (predicted) data from each class, FP and FN define miss-classified data from one class to another. Taking into account the values of average accuracies in each decision block, the VMFBE feature was used as the speech/non-speech discriminator (achieved avg. accuracy 91.47%) and the classification architecture illustrated on Fig. 6 as the music/env. sound discriminator (avg. accuracy for music/env. sound 66.10%). The most difficult task in audio-content analysis is to discriminate between voiced and unvoiced speech that interfere with music and environment sound. We have implemented the SVM classifier [8] in order to solve this task. The SVM classifier is originally designed to solve binary classification problem and discrimination of multiple classes is realized by combining several binary classifiers. Discriminative function is modeled by a linear separating hyperplane with maximal or soft margin:

$$d(\mathbf{w}, \mathbf{x}, b) = \langle \mathbf{w} \cdot \mathbf{x} \rangle + b = \sum_{i=1}^{l} w_i x_i + b, \qquad (2)$$

where l represents all the training data, $\mathbf{w}$ is a weight vector, $\mathbf{x}$ defines a support vector of training data and the scalar b is called *bias*. If the training data from different classes cannot

TABLE I
DURATION OF TRAINING AND TESTING AUDIO DATA

Audio events	Training set (min)	Testing set (min)
Pure speech	317.33	307.43
Speech with env. sound	166.68	174.36
Speech with music	100.41	106.94
Music	59.25	56.30
Env. sound	6.26	7.13
Silence	18.31	17.97
Total	**668.24**	**670.13**

TABLE II
USED FEATURES

Feature (descriptor)	*MFCCs*	*VMFBE*	*VAMFCC*	*RSF*	*BP*
Dimension	13	1	1	1	4

TABLE III
EVALUATION OF THE CLASSIFICATION ARCHITECTURES

Accuracy [%]	**Pure speech**	**Speech& music**	**Speech & env. sound**	**Music**	**Env. sound**	**Avg**
OAO-SVM	66.25	72.13	68.89	92.86	68.83	**73.79**
OAO-SVM & Rule-based	58.59	72.15	69.12	70.46	61.73	**66.41**
BDT-SVM	50.79	68.48	75.88	95.38	32.26	**64.56**

be linearly separated in the original input space, the SVM at first nonlinearly transforms the original input space into the high-dimensional *feature space*. This transformation can be achieved by using various nonlinear mappings such as: polynomial, sigmoid as in multilayer perceptrons, RBF mappings having as basic function radially symmetric function, i.e. Gaussian and spline functions.

Two classification topologies were implemented in our architecture, namely OAO-SVM and BDT-SVM. In OAO-SVM classification topology only two classes are classified at the same time. BDT-SVM architecture uses coarse-to-fine classification strategy that classifies two classes at the top of the decision tree architecture, depicted on Fig. 7. 5-fold cross-validation and RBF kernel function were used in order to set the optimal parameters of each examined SVM model. The overall accuracy of predicted data was evaluated by the same criterion as was used in the rule-based approach (1).

IV. EXPERIMENTAL RESULTS

All the evaluations were performed by using audio data obtained from the KEMT-BN1 database, which contains 188 audio recordings in PCM $16kHz$ $16bit$ mono format from the Slovak TV broadcast audio streams [9]. Total duration of audio recordings is about 65 hours. Detailed description about used training and testing set is listed in Tab. I.

Three different classification schemes were evaluated in our experiments. Tab. III shows the classification efficiency of investigated classification architectures, namely the OAO-SVM using features listed in Tab. II, the OAO-SVM utilizing rule-based and model-based approach and finally the BDT-SVM architecture. Tab. I reflects the fact that there is a big difference in amount of audio data between each class. In other words, training data are unbalanced and this fact causes the overfitting of the SVM classifier. Therefore, we implemented a simple process of data rearrangement in which the training data are arranged sequentially (one after another according to the number of classes) and form the input matrix for the classifier. The length of the input matrix (number of rows - frames) corresponds with the number of investigated classes and the length of the smallest class. The penalization parameter C was set to 16 and the parameter of the RBF function $gamma$ was set to the value 4 for each SVM model. All experiments were performed by LIBSVM software and available tools, which cooperate with this software. MATLAB software was used in order to evaluate the characteristics of features, described more in Sec. II.

V. CONCLUSION

The investigation of proposed classification architecture, applied to solve a discrimination problem in BN audio data stream, was presented in this paper. The rule-based approach in combination with the model-based classification procedure, which employs the SVM classifier, were evaluated and adjusted in order to achieve the highest classification accuracy for BN audio data. Four adopted feature parameters were involved in the process of audio events discrimination, namely VMFBE, VAMFCC, RSF, and BP. Using of these features helped us to built an optimal speech/non-speech and music/env. sound discrimination scheme as well as to create an input matrix for the SVM classifier. Results obtained from Tab. III assess the classification ability of three examined architectures. The highest classification performance was achieved by using the OAO-SVM architecture. The results also show nearly 2% improvement of the rule-based approach in combination with OAO-SVM in comparison with the BDT-SVM. Relatively poor prediction accuracy of the BDT-SVM scheme was caused by the miss-classification error that propagates from the top of the architecture and by overfitting the SVM in case of music/env. sound discrimination model (Fig. 7).

The possible way how to increase the prediction accuracy of each classification architecture is to find more appropriate features for each discrimination model and to develop much more effective classification architecture. Therefore, the area of such research will be dealt in the future work.

ACKNOWLEDGMENT

The research presented in this paper was supported by the Research and Development Operational Program funded by the ERDF under the project ITMS-26220220155 (100%).

REFERENCES

[1] W. Wang, W. Gao, and D. Ying, "A fast and robust speech/music discrimination approach," in *Proceedings of the 2003 Joint Conference on Multimedia*, vol. 3, dec. 2003, pp. 1325 – 1329 vol.3.

[2] L. Lu, H.-J. Zhang, and H. Jiang, "Content analysis for audio classification and segmentation," *Speech and Audio Processing, IEEE Transactions on*, vol. 10, no. 7, pp. 504 – 516, oct 2002.

[3] M. Kos, M. Grasic, and Z. Kacic, "Online speech/music segmentation based on the variance mean of filter bank energy," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, no. 1, p. 628570, 2009.

[4] K. Han and D. Wang, "An svm based classification approach to speech separation," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, May, pp. 4632–4635.

[5] G. Kim, Y. Lu, Y. Hu, and P. C. Loizou, "An algorithm that improves speech intelligibility in noise for normal-hearing listeners," *The Journal of the Acoustical Society of America*, vol. 126, no. 3, pp. 1486–1494, 2009.

[6] S. M. Alnadabi, Muhamad, "Speech/music discrimination: Novel features in time domain," Ph.D. dissertation, Durham University, 2010.

[7] C. K. On and P. M. Pandiyan, "Mel-frequency cepstral coefficient analysis in speech recognition," *Computing & Informatics 2006, ICOCI'06*, no. 2, pp. 2–6, 2006.

[8] S. Abe, *Support Vector Machines for Pattern Classification*, ser. Advances in Pattern Recognition. Springer-Verlang, 2005.

[9] M. Pleva, J. Juhár, and A. Čižmár, "Slovak broadcast news speech corpus for automatic speech recognition," in *Proc. of RTT '07*, 2007.

Comparison of Text Document Clustering Algorithms in Slovak

[1]*Daniel ZLACKÝ (1st year)*
Supervisor: [2]Anton ČIŽMÁR

[1,2]Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

[1]daniel.zlacky@tuke.sk, [2]anton.cizmar@tuke.sk

***Abstract*— Clustering is a technique, which organizes similar objects into clusters. Clustering can be used in many areas such as medicine, biology, marketing or computer science. In the information retrieval systems are these objects text documents. In this paper we focus on clustering Slovak text documents to specific categories using different clustering algorithms such as K-Means, K-Medoids and agglomerative hierarchical clustering using several similarity measures. We want to obtain enough knowledge, which can help us to build more robust and convenient clustering algorithm for large text corpora in Slovak.**

***Keywords*— flat clustering, hierarchical clustering, similarity measures, vector space model.**

I. Introduction

The number of documents, articles and books on the Internet is growing every day. These have brought several challenges and problems for the effective search and organization of this text data.

Clustering is a useful technique, which organizes object into smaller groups. Text document clustering groups similar documents that to form a coherent cluster, while documents that are different have separated apart into different clusters [1]. This property can be used in many various areas, where we want to organize and sort objects. In digital libraries or online stores are objects sort into several categories. If we are interested in fantasy literature, we can browse books only from this specific domain not from the all categories.

Clustering can also improve a language model for large vocabulary continuous speech recognition system in Slovak language, because we can construct domain-specific corpora from already collected and prepared Slovak text data [2] [3].

This paper is organized as follows. Section II will described how we can represent text documents, several weighting schemes and similarity measures that can be used in this area. Several basic clustering algorithms will be presented in Section III. Experiment settings, evaluation, results and analysis will be explained in Section IV. Conclusions and future work will be described in Section V.

II. Text Document Representation

Text documents are represented in the vector space model (VSM) by term vectors [4]:

$$D = (t_1, t_2, \ldots, t_n) \tag{1}$$

where t_n is a term, which contain document D. The first step in creating VSM is the text normalization. We need to remove all special markup tags, formatting, punctuation and the remaining text is parsing.

We can reduce the dimensionality of VSM by removing the stop words. There are function words (prepositions, conjunctions) and the others words with non-descriptive character for the topic of a document.

Remaining words can be stemmed by removing the different ending in one word. For example words like "*kamerou"*, "*kamerovali"*, "*kamerovými"* will be mapped to a single word "*kamer"*.

Weighting the terms is the last step in creating the VSM. There are many weighting scheme, which are suitable in our task. Term frequency (TF) model measures the frequency of occurrence of the terms in the document:

$$tf_{t,d} = \sum_{x \in d} f_t(x) \tag{2}$$

where $f_t(x)$ will be 1, if term t occurs in the document d. A relevant measure cannot only take TF into account, but a new collection-dependent factor must be introduced [4]. Inverse document frequency (IDF) performs this function. IDF is computed as follows [5]:

$$idf_{t,d} = \log\left(\frac{|D|}{df_t}\right) \tag{3}$$

where $|D|$ is the number of document in our collection and df_t is the number of documents, in which term t appears. The combination of TF and IDF scheme give us one of the most widely used weighting schemes, which is defined as a product between TF and IDF model. The weight of a term j in document i is [5]:

$$w_{i,j} = tf_{i,j} \times idf_j = tf_{i,j} \times \log\left(\frac{|D|}{df_t}\right) \tag{4}$$

A. Similarity Measures in VSM

To compute how to similar documents in our collections are, we need to use one of the proposed measures. The Euclidean distance is the basic measure, which computes the distance between two points in n-dimensional space. The distance between two documents is [1]:

$$D_E(\vec{t}_a, \vec{t}_b) = \left(\sum_{t=1}^{m} |w_{t,a} - w_{t,b}|^2\right)^{1/2} \tag{5}$$

where the documents are represented by their term vectors

$\vec{t}_a$ and $\vec{t}_b$ using TF-IDF weighting scheme. Cosine similarity computes the cosine of the angle between the two vectors. The similarity between the documents is given by [1]:

$$SIM_C(\vec{t}_a,\vec{t}_b)=\frac{\vec{t}_a\cdot\vec{t}_b}{|\vec{t}_a|\times|\vec{t}_b|} \quad (6)$$

Another measure, which can be used to compute the similarity between the documents, is Pearson correlation coefficient. It is defined as follows [1]:

$$SIM_P(\vec{t}_a,\vec{t}_b)=\frac{m\sum_{t=1}^{m} w_{t,a}\times w_{t,b}-TF_a\times TF_b}{\sqrt{\left[m\sum_{t=1}^{m} w_{t,a}^2-TF_a^2\right]\left[m\sum_{t=1}^{m} w_{t,b}^2-TF_b^2\right]}} \quad (7)$$

where $TF_a=\sum_{t=1}^{m} w_{t,a}$ and $TF_b=\sum_{t=1}^{m} w_{t,b}$.

There exist a lot of others similarity measures, which can be used in computing similarity between documents such as Jaccard coefficient, Dice coefficient, Kullback-Leibler (KL) divergence and the others. In [1] author compare similarity measures in text document clustering. The result was that the Pearson correlation coefficient and the averaged KL divergence measures were better than others.

III. Clustering Algorithms

This section focuses on clustering methods and explains several basic algorithms. Clustering is a technique, which classify a set of objects into groups or clusters based on their similarity. The goal is to create clusters that are coherent internally, but clearly different from each other. Documents in the same cluster should be as similar as possible to each other and the cluster should be as dissimilar as possible to another cluster [6]. Clustering algorithms can be classified into a two basic types: *hierarchical clustering* and *flat clustering*.

A. Hierarchical Clustering

Hierarchical clustering creates a hierarchy of clusters [6]. It can be also classified into a two basic types. Agglomerative clustering provides a "bottom up" clustering. Each object is in its own cluster at the beginning. It merged the most similar pairs of clusters and ends, when all objects are in the one cluster. Divisive technique provides a "top down" clustering. It starts with each object in the one cluster and splits them to smaller clusters based on their similarity.

Similarity between the documents is computed by one of the measures proposed and described in the Section II. Next, it is necessary to specify, which similarity function will be used for merging the clusters. The single-linkage computes the distance between the most similar members in two clusters. The distance between the clusters *P* and *Q* will be:

$$D(P,Q)=\min d(p,q) \quad (8)$$

where $p\in P$ and $q\in Q$. The complete-linkage computes the distance between the most dissimilar members:

$$D(P,Q)=\max d(p,q) \quad (9)$$

Group-average similarity function is a compromise between single-linkage and complete-linkage. The criterion for merging is the average similarity between clusters [7]:

$$D(P,Q)=\frac{1}{|P|\cdot|Q|}\sum_{p\in P}\sum_{q\in Q} d(p,q) \quad (10)$$

There exist several others similarity function like Ward`s criterion. In [8] the best clustering result were achieved with Ward`s and group-average criterion.

B. Flat Clustering

Non-hierarchical clustering algorithms consist of a certain number of clusters and the relation between them is often undetermined. Most of these algorithms are iterative. They provide a reallocation operation that reassigns objects [7].

K-means is one of the simplest iterative clustering methods. It works in these steps [9]:

- Choose k objects as the initial cluster centers.
- Repeat
  - (Re)assign each object to the cluster based on the given similarity function
  - Update the centroid
- Until no change

The centroid of the cluster C_k is defined as [10]:

$$m^{(k)}=\left(\frac{1}{n_k}\right)\sum_{i=1}^{n_k} x_i^{(k)} \quad (11)$$

In the Figure 1 are above steps graphically illustrated. Two cluster centers are set at the first step. All documents are assigned to one of these centers. Then the centers are recomputed and moved to another location. The convergence is reached after the nine iteration.

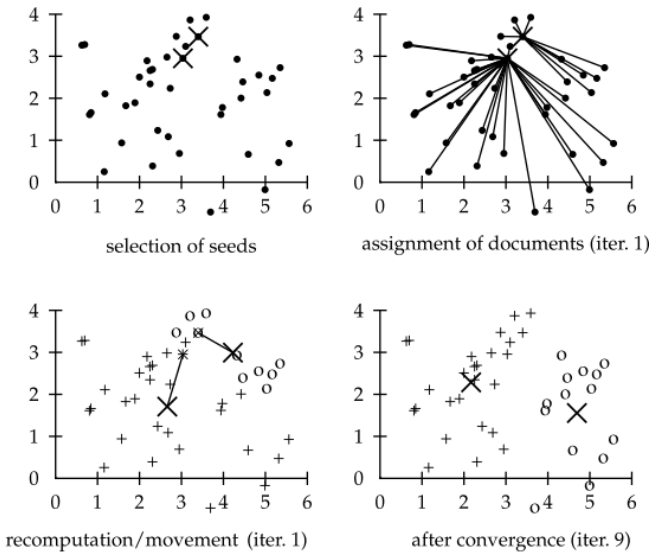


Fig. 1. K-Means example [4]

It works well, when we are processing large data sets and it is very easy to implement, but it is sensitive to outliers.

K-medoids clustering algorithm uses medoids to represent the clusters. A medoids is the most centrally located object in the cluster [9]. K-medoids algorithm is more robust than K-means, it is less sensitive to outliers, but the implementation is much more complicated and is suitable, when we are processing smaller data sets.

C. Evaluation of Clustering

Purity and entropy are basic evaluation measures, which are used to evaluate the quality of different clustering techniques. Purity is defined as the proportion between the numbers of documents from a single category to the total number of documents in the given cluster. The purity of a cluster C_r is

defined [11]:

$$P(C_r) = \frac{1}{n_r} \max_i n_r^i \tag{12}$$

where $\max_i(n_r^i)$ is the number of documents from the dominant category and n_r is the size of the particular cluster. The purity of the entire clustering solution is the weighted sum of the individual cluster purities [11]:

$$purity = \sum_{r=1}^{k} \frac{n_r}{n} P(C_r) \tag{13}$$

A perfect clustering solution is, when the cluster contains documents only from the single category with purity close to 1.

Entropy evaluates how different documents are distributed within each cluster. The entropy of cluster C_r is defined [11]:

$$E(C_r) = -\frac{1}{\log q} \sum_{i=1}^{q} \frac{n_r^i}{n_r} \log \frac{n_r^i}{n_r} \tag{14}$$

where q is the number of categories in the corpora and n_r^i is the number of documents from i-th class that were assigned to the r-th cluster. The entropy of the entire clustering solution is the weighted sum of the individual cluster entropies [11]:

$$entropy = \sum_{r=1}^{k} \frac{n_r}{n} E(C_r) \tag{15}$$

The clustering solution is better, when the entropy values are close or equal to 0.

IV. Results

Our database consists from 935 Slovak articles gathered from Wikipedia, which were manually sorted into 8 categories. These documents were stemmed using unsupervised morpheme segmentation with Morfessor [13]. We created list of 648 Slovak stop words, which were removed from these documents. In the Table I. are illustrated the statistics of the created database. The most documents were from Music category and the least from Cars category.

TABLE I
Statistics of the Created Database

Category	*Documents*	*Unique words*
Cars	59	4 069
Films	119	7 529
Hockey	76	3 085
Music	280	16 690
Medicine	113	10 572
Computers	81	6 889
Law	100	7 109
Space	107	8 579
Total	935	64 522

The TF-IDF model was chosen as the weighting scheme. In the Table II are summarized several basic statistics of it. 935 text documents consist from 84 478 unique features. Maximum number of feature in one document were 1 115, with average number of 459 words. Maximum occurrence of same feature in one document was 97, with average number of 7.86. All features, which occurrence was less than 3, were removed from this model. The final count of features were 19 460 after reduction.

TABLE II
Statistics of the TF-IDF Model

N-gram	*Unigram*
Count of documents	935
Count of features	84 478
Max. number of features in document	1 115
Avg. number of features in document	459
Max. frequency of features in document	97
Avg. frequency of features in document	7.86
Max. TF-IDF of a feature	2.75
Avg. TF-IDF of a feature	0.06
Min. occurrence of a feature	3
Final count of features	19 460

We tested 3 basic unsupervised clustering algorithms and several similarity measures and functions, included in the R project for statistical computing. Pearson and Correlation similarity measures were used in K-Means algorithm in the library *amap* using function *kmean*. Spearman and Euclidean similarity measures were used in K-Medoids algorithm in the library *cluster* using functions *pam* and *clara*. The similarity measure for hierarchical clustering was set to Pearson using Group-Average similarity function in the library *stats*.

The purity results for 3 basic unsupervised clustering algorithms are illustrated in Table III. The best result was achieved with hierarchically-based clustering, where 99% documents were assigned to the right category. The worst result achieved K-Medoids clustering with Spearman similarity measure, where only 79% documents were assigned to the right category.

The entropy results are illustrated in the Table IV. The results are similar to the purity results. Hierarchical clustering achieved the best results with weighted sum of the individual cluster entropy close to 0. K-Means and K-Medoids algorithm with Euclidean distance measure also achieved good results.

V. Conclusion

In this paper we compared several clustering algorithms and similarity measures. The best result was achieved by the hierarchical clustering with group-average similarity function. Good results also achieved both K-means similarity measures and K-Medoids with Euclidean distance as similarity measure. The worst result achieved K-Medoids with Spearman correlation coefficient as similarity measure.

We proved that the clustering is useful technique, if we want to organize text documents. We achieved very good results, but our database was very small. It's uncertain, what results will be achieved with bigger database, which will consist from thousands documents.

Our future work will focus on creating clustering algorithm based on keywords. We want to extract keywords from certain category such as law or medicine and use these keywords to cluster other text document.

TABLE III
PURITY RESULTS

Algorithm	*K-Means*		*K-Medoids*		*Hierarchical*
Measure	*Pear.*	*Corr.*	*Spear.*	*Euclid.*	*Average*
Cars	0.89	0.92	0.88	0.94	1.00
Films	0.99	0.99	1.00	0.99	1.00
Hockey	0.99	0.99	1.00	0.96	1.00
Music	0.99	0.99	0.86	0.99	0.99
Medicine	0.98	0.94	0.82	0.99	0.99
Computers	0.66	0.54	0.77	0.97	0.98
Law	0.96	0.95	0.89	0.97	0.98
Space	0.99	0.98	0.85	0.96	0.99
Weigh. avg,.	0.88	0.85	0.79	0.94	0.99

TABLE IV
ENTROPY RESULTS

Algorithm	*K-Means*		*K-Medoids*		*Hierarchical*
Measure	*Pear.*	*Corr.*	*Spear.*	*Euclid.*	*Average*
Cars	0.20	0.16	0.23	0.07	0.00
Films	0.02	0.02	0.00	0.02	0.00
Hockey	0.03	0.03	0.00	0.06	0.00
Music	0.01	0.01	0.22	0.02	0.04
Medicine	0.04	0.10	0.31	0.06	0.02
Computers	0.46	0.43	0.44	0.12	0.06
Law	0.10	0.11	0.22	0.08	0.05
Space	0.03	0.05	0.28	0.06	0.03
Weigh. avg.	0.09	0.09	0.18	0.06	0.03

ACKNOWLEDGMENT

The research presented in this paper was supported by the Ministry of Education, Science, Research and Sport of the Slovak Republic under research projects No.3928/2010-11 (50%) and by Research & Development Operational Program funded by the ERDF under the ITMS project 26220220155 (50%).

REFERENCES

[1] A. Huang. "Similarity measures for text document clustering," *Proceedings of the Sixth New Zealand Computer Science Research Students Conference*, pp 49-56, 2008.

[2] J. Juhár, J. Staš, D. Hládek. "Recent progress in development of language model for Slovak large vocabulary continuous speech recognition," *New Technologies: Trends, Innovations and Research*, Rijeka: InTech, p. 261 – 276, 2012.

[3] D. Hládek, J. Staš. "Text mining and processing for corpora creation in Slovak language," *Journal of Electrical and Electronics Engineering*, vol. 3, no. 1, p. 65 – 68, 2010.

[4] G. Salton, Ch. Buckley. "Term-weighting approaches in automatic text retrieval," *Information Processing and Management* Vol. 24, No. 5, pp 513-523, 1988.

[5] D. L. Lee, H. Chuang, K. E. Seamons. "Document ranking and the vector-space model," *Software, IEEE*, vol. 14, no. 2, pp 67-75, 1997.

[6] Ch. D. Manning, P. Raghavan, H. Schütze. *An Introduction to Information Retrieval*. Cambridge University Press, 1 edition, 2008.

[7] Ch. D. Manning, H. Schütze. *Foundations of Statistical Natural Language Processing*. The MIT Press, 1 edition, 1999.

[8] L. Ferreira, D. B. Hitchcock. "A comparasion of hierarchical methods for clustering functional data." *Comm Stat Simulation Computation*, pp 1925-1949, 2009

[9] S. S. Singh, N. C. Chauhan. "K-means v/s K-medoids: A comparative study." *National Conference on Recent Trends in Engineering & Technology*, Gujarat, India, 2011.

[10] A. K. Jain, R. C. Dubes. *Algorithms for Clustering Data*. Prentice Hall, 1988.

[11] T. Van de Cruys. *Mining for Meaning: The Extraction of Lexico-semantic Knowledge from Text*. Dissertation thesis, Rijksuniversiteit Groningen, Netherlands, 2010.

[12] R. Kohavi. "A study of cross-validation and bootstrap for accuracy sstimation and model selection." *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, vol. 2, pp 1137-1143, 1995.

[13] J. Staš, D. Hládek, J. Juhár, D. Zlacký. "Analysis of morph-based language modeling and speech recognition in Slovak," *Advances in Electrical and Electronic Engineering*, vol. 10, no. 4, p. 291 – 296, 2012.

Control of OWI 535 robotic arm using C# and possibilities of future applications

[1]*Jakub ČERKALA (1st year)*
Supervisor : [2]Anna JADLOVSKA

[1] Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

[1]jakub.cerkala@tuke.sk, [2]anna.jadlovska@tuke.sk

***Abstract*— In this article will be presented the control of robotics arm model OWI 535 using USB interface. Control program is designed as C# library functions and its goal is the practical use of simple robotics arm model for educational purposes. Within this article will be presented and described robotics arm, its parameters and kinematics model. Furthermore, direct communication using USB between robot's control unit and own application with graphical user interface will be described as well. In addition, the theoretical and practical options of model usage in education will be also presented.**

***Keywords*— robotic arm, C# programing language, graphical user interface, .NET library, communication via USB**

I. Introduction

Robotics and mainly robotic arms represent nowadays a wide field of science and they are often combined with mobile robotics branch and artificial intelligence methods usage [1][2][3]. There are several types of robotic arms, mainly divided by the number of degrees of freedom and use.

Robotic arms are designed mainly for industrial use and they are widely used in manufacturing, usually with robust design to allow one type of robot to be used for variety of purposes just by changing the software and tool, which is placed in the wrist of the robot. In most cases, as joints are driven by powerful servos and for this reason they are relatively expensive and less available.

This article goal is to present the possibilities of enhancement and usage of simple robotics arm for educational purposes. This paper deals with the OWI 535 robotics arm model with USB control unit, shown on Fig. 1.

II. Robotic arm model description

A. Device overview

OWI 535 itself offers an affordable robotic arm with 4 +1 degrees of freedom, which is designed for educational purposes [4]. There are various versions of mentioned model, this article describes the version, for which the robot is connected to the PC using the USB interface. The specified robotic arm model is supplied with driver for x86 and x64 operating systems and simple application for control, but this application is only an executable application without source code, which can't be modified.

Fig. 1. Maplin OWI 4 +1 DOF robotic arm with USB interface.

B. Properties and motion ranges

As mentioned earlier, the robotic arm has 4 +1 degrees of freedom, and they can be represented by subsystems: base rotation, base angle, elbow, wrist and gripper. Ranges of possible motion are listed in Table I. Robot uses independent 3V power supply from batteries and control unit is powered via USB interface. In present, model's power supply has been replaced by common PC stabilized power source to improve of motion characteristics, since batteries cannot guarantee stabile current ant voltage during whole time in use.

TABLE I
Subsystem Ranges

Subsystem	Name	Range
S1	Base rotation[1]	270 degrees
S2	Base angle	180 degrees
S3	Elbow joint	300 degrees
S4.	Wrist joint	120 degrees
S5	Gripper	Approx. 0.045 m wide grip

[1] Base can be mounted two different ways – desired rotation from control unit can be chosen.

Subsystem contains DC engines and gearboxes, which is the main weakness of the model in comparison to robots used in industry and medicine. However, gearboxes are constructed safeguard against over spin and therefore the damage resistance and durability is higher. Robotic arm uses identical gearboxes for all joints. A LED lamp is mounted into gripper, and it can be controlled like other subsystems. Robot can lift

approximately 100g. Vertical reach is approx. 380 mm and horizontal reach is approx. 320 mm.

C. *Kinematic model of robotic arm*

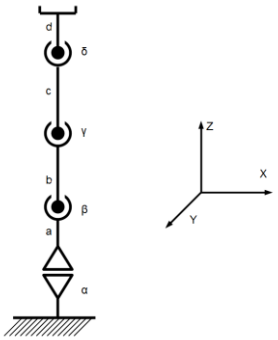


Fig. 2. Robotic arm kinematic model.

It is assumed that the robotic arm is rotated in the middle of base rotation position and the rest of the joints are rotated up and thus the arm is in maximum vertical position. After obtaining the dimensions and distances between joints, a kinematic model of the robotic arm can be built. With use of the matrixes R_Y, R_Z for rotational motion around *Y* axis (1) and *Z* axis (2) a direct kinematic model can be obtained [5], shown on Fig. 2., according to which gripper position in space for specific base rotation α and joint rotations β, γ a δ can be computed.

$$R_Y(\theta):\begin{pmatrix}\cos\theta & 0 & \sin\theta\\ 0 & 1 & 0\\ -\sin\theta & 0 & \cos\theta\end{pmatrix}, \tag{1}$$

$$R_Z(\theta):\begin{pmatrix}\cos\theta & -\sin\theta & 0\\ \sin\theta & \cos\theta & 0\\ 0 & 0 & 0\end{pmatrix}. \tag{2}$$

To full use of this advantage, model needs to be enhanced by robotic arm positional sensors, which can be realized as optocouplers, connected to lab card [6].

An alternative to optocouplers is the application of the potentiometers, but their installation is more complicated and they require an analogue lab card. In addition to sensing of the joint positions, a black and white camera connected to a PC through a USB interface can be added, when placed on the gripper it can increase overall possibilities of the model.

III. Communication Subsystem

To be able to move the robot each engine needs to be actuated with voltage in form of correct bites pair. To do such task, robotic arm control unit can be used, because it is connected to PC via USB interface. As programming language was chosen the object oriented C# language [7].

An open source library *LibUsbDotNet.dll* is used to control the device by USB interface. There was a problem with electronically unsigned drivers supplied with robotic arm - for x64 architecture operating system, but this library already includes the ability to work with signed or unsigned drivers and own drivers as well. Such drivers have to be manually added by user in a special application that also allows the user to view the details list of a device connected to the USB. Robotic hand can be connected and selected only if it's the *Vid* - vendor ID and *Pid* - Product ID are known. Those two numbers are used to identify the any device and can be found in the above-mentioned details list.

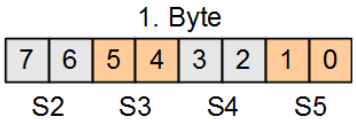


Fig. 3. Control byte pairs scheme for S2-S4 robot joint control.

Robot receives 3 control bytes in a row. The first control byte carries information for changing the orientation and movement of joints S2 to S4, specific pairs are shown in Fig. 3. The second byte is for control of the base rotation and last manages only LED lamp state on gripper. Control byte setting is based on specific bit pairs for selected subsystem, there are three possible positions – 00 means always stop, 01 is for positive rotation direction and 10 is for negative direction. LED lamp is controlled only by one bit, 0 means that the lamp is off.

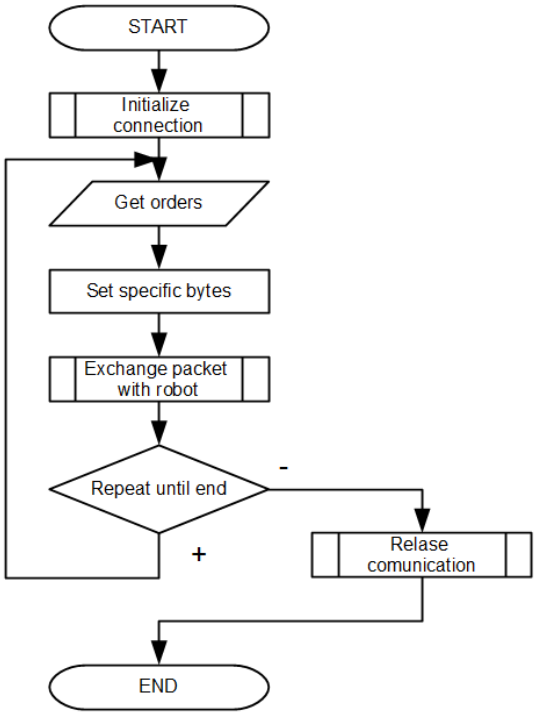


Fig. 4. Algorithm describing asynchronous communication between PC and robotic arm

Methods of created library enable sending a packet that consists of mentioned 3 bytes. A synchronous sending option sending data packets in periodic intervals was also used, but it only uses the methods for asynchronous packet sending.

Before the packet can be sent, a check if the device is connected to USB port needs to be done. Program browse the list of all connected devices and look up for robotic arm based on Vid a Pid. If such device is present in the list, the robotic arm is presumed as connected and it can receive control

packets. This initialization process can be summed into block *Initialize Connection* in state-flow diagram show on Fig. 4. This state-flow diagram describes the algorithm of communication loop, namely get orders for joint rotation, set specific bits in control byte of packet and the block *Exchange packet with robot* represents sending the bytes in a row. A graphical user interface (GUI) as Windows Forms application, shown on Fig. 5, was designed for easy model use.

Fig. 5. Own graphical user interface.

IV. Graphical User Interface

GUI, that was designed allow asynchronous control of one robotic arm with commands for each subsystem. Information about subsystem state is displayed in table and status bar contains actual packet bit setting for each byte. For safety reason, there is also included a Total stop mode, which after activation immediately stops all subsystem movement and blocks sending of any next packets. Main purpose of this GUI is ability to direct control of robotic arm model by user and calibration tasks.

V. Experiments and Tests

Since this robot does not contain components that enable feedback control that are intended to be added in near future, experimental options are limited. An experiment of base and joint rotation was conducted, where the angular deviation of joint for specific time was measured. By this way it was able to determine exact time in milliseconds needed to move joint for desired angle. This experiment also revealed that the direction of movement change need to be compensated with extra time and can be counted how much time is needed. Linear characteristics for angular changes larger than 10 degrees were measured. However, the precision of this kind of measurements is insufficient and results were affected by battery power supply. This experiment will be repeated after adding the optocouplers and with stabilized power.

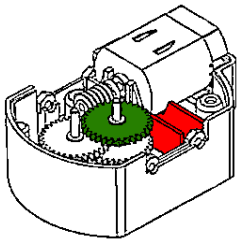


Fig. 6. Possible position of optical counters in robotic arm gearbox.

Optocouplers will be placed into gearboxes, to highlighted position shown on Fig. 6. and they will count the rotations of topmost gear, which is driven directly by DC engine. Gear will have six holes, which will suffice for desired accuracy. The rotations of this gear than can be converted into base and joint angular deviation.

VI. Conclusion

Despite high prices and limited offer, it is today possible to obtain a model of the robotic arm, which is, although it is not an industrial version, an affordable device for testing purposes of simple algorithms, and the general laws of physics. The aim of presented model is to use is options. The goal of this paper is to utilize the possibilities of the mentioned model for verification of complex control algorithms that use artificial intelligence techniques. However, before this algorithm can be practically tested, the model needs to be improved by adding sensors that enrich its options.

If the position sensors will be installed, robot could be used in trajectory tracking by direct and inverse task with use of its kinematic model [8]. As mentioned earlier, robot should have a camera mounted on gripper. It could be used for face or target tracking with use of computer vision methods. Neural networks can be used in motion and pattern recognition [9] [10].

Acknowledgment

This work has been supported by the Scientific Grant Agency of Slovak Republic under project Vega No.1/0286/11 Dynamic Hybrid Architectures of the Multiagent Network Control Systems (50%). This work is also the result of the project KEGA 021TUKE-4/2012 (50%).

References

[1] Mobile robot Khepera II - simulation approach. Center of Modern Control Techniques and Industrial Informatics [online]. [cit. 2013-03-11]. Available from: http://kyb.fei.tuke.sk/laboratoria/modely/khe.php

[2] ŠUSTER, P., JADLOVSKÁ, A.: Tracking trajectory of the mobile robot Khepera II using approaches of artificial intelligence / 2011.In: Acta Electrotechnica et Informatica. - ISSN 1335-8243. - Yr. 11, č. 1 (2011), pages 38-43.

[3] MASÁR, I. Inteligentný regulátor na sledovanie trajektórie mobilným robotom. Automatizace [online]. 2007, Yr. 50, num. 2, s. 80-85 [cit. 2013-03-10]. ISSN 0005-125X. Available from: http://www.automatizace.cz/article.php?a=1629

[4] MAPLIN ELECTRONICS. Robotic arm with USB pc interface: Assembly and Instruction manual [online]. 29 pages. [cit. 1.3.2013]. Dostupné z: http://www.maplin.co.uk/media/pdfs/A37JN_manual.pdf

[5] KROKAVEC, D. Spracovanie údajov v robotike. 1. ed. Košice: Rektorát Vysokej školy technickej v Košiciach, 1985. ISBN 85-629-85.

[6] MF624: Manuál. HUMUSOFT. [online]. [cit. 2013-03-11]. Available from: http://www.humusoft.cz/produkty/datacq/mf624/

[7] SHARP, J. Microsoft Visual C# 2008: Krok za krokem. 1. ed. Brno: Computer press, a. s., 2008, 592 s. ISBN 978-80-251-2027-9.

[8] PASTOREKOVÁ, L., MÉSZÁROS, A. BURIAN, P. Inteligentné riadenie systémov na báze inverzných neuronových modelov. AT& P PLUS. 2005, č. 7, pages. 53-58. ISSN 1335-2237.

[9] PHAM, D., LIU X. Neural networks for identification, prediction and control. 2. vyd. London: Springer-Verlag, 1995. ISBN 3-540-19959-4.

[10] KAJAN, S. Comparison of some neural control structures for nonlinear systems. Journal of Cybernetics and Informatics [online]. 2009, 11 pages [cit. 2013-03-10]. ISSN 1336-4774. Available from: http://kasr.elf.stuba.sk/sski/casopis/show_article.php?id_art=55&id_cat=9

Language Enrichment Environment: Proposal and Experiments

[1]*Emília PIETRIKOVÁ (3^{rd} year),* [2]*Sergej CHODAREV*
Supervisor: [1]*Ján KOLLÁR*

[1,2]Department of Computers and Informatics, FEI TU of Košice, Slovak Republic

[1]emilia.pietrikova@tuke.sk, [2]sergej.chodarev@tuke.sk

***Abstract*—This article is focused on presentation of a method for automated raise of programming language abstraction level. The base concept is a code pattern – recurring structure in program code. In contrast to design patterns it has a specific representation at a code level and thus can be parameterized and replaced by a new language element. The paper presents an approach for language extension based on the found patterns. It is based on an interactive communication with the programming environment, where recognized patterns are suggested to a programmer and can be injected into the language in a form of new elements. An algorithm for automated recognition of patterns is described and examined as well. Conducted experiments are evaluated in regard to the future perspective and contributions.**

***Keywords*—Abstraction, code patterns, language extension, projectional editing.**

I. Introduction

With growth of software systems, expression complexity of their properties in a programming language mounts up as well. As the answer to complexity, higher levels of abstraction can be introduced, encapsulating complex expressions [1], allowing hiding the implementation details. A promising solution can be an abstraction based on a language, allowing reduction of the complexity through defining new, more abstract concepts and language constructions.

As a background of abstraction we regard Voelter's *DSL Engineering* book [2] stating 2 ways of defining abstractions:

1) *Linguistic*, with abstractions built into the language, what is typical for DSLs (Domain-Specific Languages),
2) *In-language*, with abstractions expressed by concepts available in the language, typical for GPLs (General-Purpose Languages).

Linguistic abstraction assumes a programmer is able to use correct abstraction since DSL users are familiar with a particular domain. Thus, no analysis to reverse engineer the semantics is needed and no details irrelevant to the model are expressed, increasing conciseness and avoiding over-specification [2]. Through transformation rules tied to the identities of the language concepts, linguistic abstraction represents a simple process making the language suitable for domain experts. However, this way the language evolution has to be frequent and the abstractions have to be known in advance to be used.

Compared to the linguistic, in-language abstraction constitutes achieving conciseness by a GPL providing facilities allowing users to define new (non-linguistic) abstractions in programs [2], including procedures or functions, and higher-order functions or monads. This way abstraction can be provided, but no declarativeness. Thus, in-language abstraction is more flexible as users can build only those abstractions they actually need. However, while programmers are actually trained to build their own abstractions, domain experts are not.

Linguistic abstraction is a basic element of Language-Oriented Programming [3], [4]. In this methodology, the first step of program design is a definition of high-level domain-specific language suitable for solving a specific problem[1]. Next, the program itself is implemented using the new language which is built upon the existing (less abstract) language. From this point of view, each level of abstraction is represented by a language, where each is defined using a lower level one.

II. Motivation

Let us consider two pieces of pseudo-code expressing transformation of the array values:

```
results = new Array();
for (int i = 0; i < data.size; i++){
  a = data[i];
  results[i] = f(a) * 5 + 3;
}
squares = new Array();
for (int i = 0; i < numbers.size; i++){
  b = numbers[i];
  squares[i] = pow(b, 2);
}
```

As both pieces of pseudo-code are very similar, replacement of the repeated structures might be convenient. First, let us consider a new pseudo-code, applicable to both examples:

```
«output» = new Array();
for (int i = 0; i < «input».size; i++){
  x = «input»[i];
  «output»[i] = «op x»;
}
```

Where:

- «`input`» can be considered as a variable replaceable by arrays *data* and *numbers*;
- «`output`» represents a variable for arrays *results* and *squares*; and
- «`op x`» represents a variable for the calculations: $f(x)\times 5+3$ and x^2.

[1]The process of solving a problem by designing a new language first is itself also called metalinguistic abstraction [5]

Since it is possible to apply the new pseudo-code to both examples, it can be regarded as a *pattern*. Abstraction has one simple goal in mind: To replace repeated code structures in order to increase expression abilities of the language. For the discussed examples, the identified pattern might be reduced and simplified with a new construct *map* (inspired by functional programming):

```
«output» = map («input», «op x»);
```

Where *map* can be considered as an abstraction to the identified pattern, representing the whole structure of the cycle with the appropriate parameters. For the two examples, it is now possible to use new, more abstract pseudo-code (with notation backslash denoting an anonymous function). This approach enables the program code to be much shorter, thus less prone to errors:

```
output = map(input, (\x -> f(x) * 5 + 3));
squares = map(numbers, (\x -> pow(x, 2));
```

Several implications arise from the above considerations:

- If it is possible to recognize language structures within a source code, then it is possible to identify recurring structures as well.
- If there is a large group of source code belonging to the same application domain, then it is possible to identify plenty of recurring structures within the domain.
- If frequently repeated structures are abstracted into the new ones, then it is feasible to form a new language dialect.
- If the new language structures are named by concepts of the appropriate application domain, then the resultant dialect is domain-specific.
- If a programmer is able to write short codes in concepts of the appropriate application domain instead of long codes in concepts of the general-purpose language, then his work might become much more effective.
- If the programming environment would provide help with definition of new abstractions, then there is a higher chance that abstractions would be actually used.

Moreover, analysis of the current state within application of programming languages proved that along with system development in various application areas, there is a demand for the following language features [6], [7]:

- Increasing level of abstraction when expressing complex issues
- Increasing expression ability of a language, and thus effectiveness of its application
- Specialization of languages on specific domains of use
- Increasing flexibility when using a language in other domains

Considering importance of the abstraction concept in programming, there are a lot of open questions remaining, particularly regarding automatic analysis and introduction of abstraction. Therefore, in the following sections we will try to find answer (or more answers) to the following main question: *How can increase of abstraction be automated?*

Our approach to this task is based on resolving two basic problems concerning tool support for automated program abstraction and language enrichment:

1) Recognizing recurring patterns in program code.
2) Finding a way to inject identified patterns into a language as new constructs.

III. Proposal

We decided to base our approach to program abstraction on the concept of *patterns* – recurring structures in programs. The conceptual scheme of the proposal is depicted in Fig. 1.

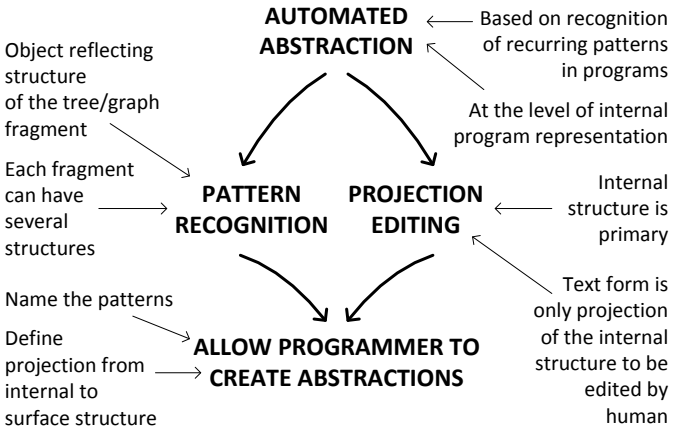


Fig. 1: Proposal conceptual scheme

To propose a solution for automated introduction of new language abstractions based on patterns found in source code the problem of recurring pattern recognition should be resolved. Manual analysis of code may be a hard and tedious task. However, a tool for automatic pattern recognition can greatly help in this task.

A. Pattern recognition

A *pattern* for our purposes is a recurring structure in program code. These structures can be expressed by fragment of a program with parts that may be different, replaced by *pattern variables*. This concept is different from *design patterns* [8] that describe recurring patterns and their usage on higher level. On the other hand, we are currently concentrating on patterns of a smaller scale.

Patterns would be recognized at the level of language abstract syntax. The abstract syntax tree or graph of a program contains all important information about structure of program code without syntactic details.

We have developed two approaches to this task:

1) Pattern recognition by comparing
2) Pattern recognition by collecting

Recognition by comparing is based on traversing trees representing programs from leaves up and comparing the subtrees to find groups of subtrees with the same structure.

The second approach that we propose is based on collection of abstracted structural schemas of program fragments. It can be obtained by replacing parts of the fragment by variables. Each fragment of code may be described by several structures of different level of detail. If such structural schemas are identified and stored with references to program fragments that contain them, it is possible to analyze frequency of their usage and by this way identify possible pattern.

B. Language enrichment

Since recognized patterns represent recurring structures found in programs, they also present potential extensions of the language. To inject a pattern into the language there is a

need to name it and define its syntactic or surface structure that would be used to represent it in code of a program.

Enrichment of a language syntax and semantics directly by its users, though, is rarely allowed. In most cases it requires modification of the original language implementation, since a composition of the language with new elements is needed. Doing so with traditional textual language processed using a parser usually generated by some parser generator according to the grammar specification, it may result in several problems caused by possible ambiguity of the resulting grammar. Partially, this is caused by the fact, that grammar subclasses supported by common parser generators are not closed under composition [9]. Moreover, to allow composition and extension of a language without modification of the original implementation, it requires special tool support [10] and knowledge in the field of language development.

A possible way to solve the problem of language composition is a transition to *concept composition*. This means that instead of composing languages and their grammar rules, only concepts in a single base language are composed. This requires lowering the role of language grammar and is possible to be achieved at least in two ways:

1) Using single syntax for composed languages.
2) Using projectional editing.

In our case, the first way means not to use special syntax for injected patterns. Patterns would only be named and one of the shapes predefined in the language syntax would be assigned to them. This is similar to the definition of a function – it is assigned a name and standard syntax of function call or operator application. Another example is extension of Lisp and its dialects using macros that are based on the uniform syntax of S-expressions [11], or definition of XML based language. Disadvantage of such an approach is indeed low flexibility of choosing an appropriate notation.

Projectional editing on the other hand keeps different notations for languages on the surface, while using unified representation internally. The syntax becomes only a matter of projection and actual information of a program (code) is stored in some different form, not visible for a language user.

Projectional editing is used by some language workbenches, for example MPS [4] or Intentional Workbench [12]. They use internal graph-based representation as main form of a program. Editable form is only a projection of internal form [13]. When a user is issuing editing commands at the projection, the internal structure is modified and the projection is updated accordingly. This allows different types of editable representation besides the textual, e.g. graphical or table-based.

In this way, elements of languages developed using a language workbench are actually only concepts of an internal representation language (which is usually not textual). This means that in this case the composition of languages corresponds to the composition of concepts inside a single language. Textual composition is only its projection which is not required to be unambiguous.

C. Concept of language enrichment environment

Based on the considerations described above, it is possible to construct an environment for automated language enrichment based on patterns found in programs. The principles of its functionality are depicted in Fig. 2. The primary representation of the program is its abstract syntactic graph (ASG) editable through the projection. Syntax specification is used as a basis for projection and editing environment.

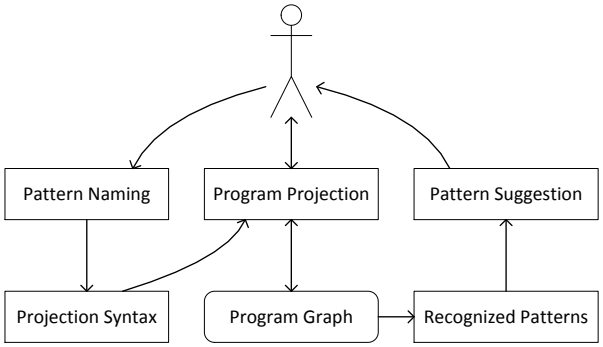


Fig. 2: Architecture of language enrichment environment

While program is created by a programmer, the structure of its elements is collected inside the environment to recognize the recurring patterns. Then, the environment should use the collection to suggest recognized patterns to the programmer. Patterns can have various uses beside the automated abstraction. For example, they can provide automatic completion of snippets for frequently used constructs.

If a programmer decides that the pattern is semantically significant, he can name the pattern and therefore enrich the language he uses. He needs to provide specific surface syntax for the named pattern that would be appropriate for conveying its meaning. A new syntax rule would be added to the language syntax definition and used by the projectional editor.

At the same time named pattern becomes a part of the language structure and semantics definition. By default it expands the pattern during program evaluation or translation. Definition though can be amended by some optimization rules based on semantics of the pattern.

Introduction of new abstractions into the language can also have a negative effect. The more new constructs specific to a program are added into the language, the more is a programmer forced to learn new abstractions. Moreover, if more than one programmer is working with the same program code, and only few of them know the new abstractions, a problem may occur as they may not understand program codes of each other. This situation may occur already in the current abstraction range of various general-purpose and domain-specific languages.

Considering this, it would be appropriate if the environment would also provide reverse mode of work. The user should be allowed to switch the level of abstraction used in the code and display it in expanded form. That is, our experiments and algorithms for pattern recognition should result in two instruments available to a user of the programming environment:

- Pattern encapsulation – pattern replacement by new syntactic element
- Pattern expansion – syntactic element replacement by its implementation via elements of lower level abstraction

For instance, if one programmer knows list comprehension construct, the other one, who does not, should be able to specify directly within the environment that he does not want to use list comprehension, or that he wants to learn their structure. Then, any list comprehension would be equivalently substituted, e.g. through map or filter function. This expansion can be even applied repeatedly as in the following sequence:

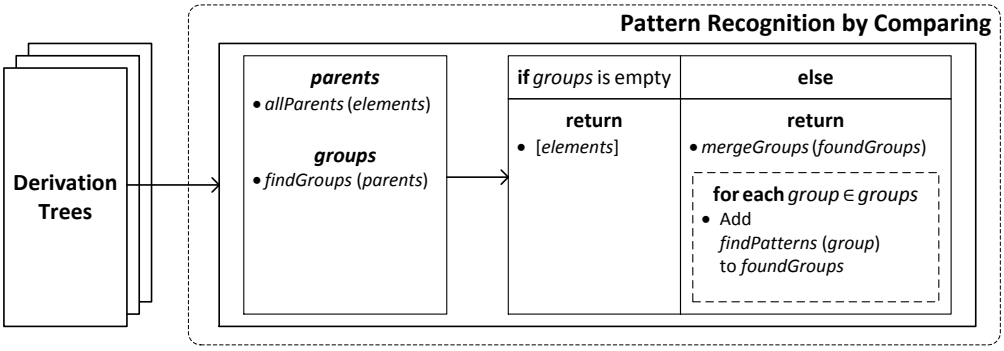


Fig. 3: Architecture of pattern recognition by comparing tools

```
squares = [x ^ 2 | x <- xs]
   ↓
squares = map (\x -> x ^ 2) xs
   ↓
squares [] = []
squares (x:xs) = (x ^ 2) : squares xs
   ↓
squares xs | null xs = []
           | otherwise = (head x ^ 2) :
                         squares (tail xs)
```

On the contrary, a programmer may also use several complex structures, and then replace them through the known patterns by equivalent, shorter structures, and thus noticeably reduce the program code length.

IV. Experiments

For pattern recognition, we suggest two different methods: by comparing and by collecting. For experimental purposes, the first method has been performed and examined on a large group of Haskell programs while the second method has been performed on a simple language of functions and expressions.

Pattern recognition by comparing is also a successor to another research, as it uses implementations of experiments associated with effects of abstraction [14]. On the other hand, pattern recognition by collecting is based on the evaluation of the first method, reducing unnecessary implementations and extending possibilities of the pattern recognition.

A. Pattern recognition by comparing

Pattern recognition method based on comparing program fragments was developed on top of the set of tools gathering information from Haskell programs to get a proper knowledge about the used constructs in analyzed programs. Its principle lies in comparing different program fragments to find groups of similar ones. To recognize syntactic patterns, it is important to decide which parts of the analyzed programs may be considered similar. Trees can be considered similar if their structure is equal except for the attributes of terminal symbols.

Since this approach does not fall within our latest experiments, we do not enclose more detailed description of the algorithm and experiment outputs. In case of an interest, it is possible to find them in [15] and [16] or see Fig. 3.

Another approach is to allow differences in entire subtrees rooted in the same type node. This should allow more complex syntactic variables and is, however, more difficult to implement using specified approach, as it requires comparisons of program fragments on different levels of the tree.

B. Pattern recognition by collecting

Another experiment has been performed to evaluate different algorithm for pattern recognition. It is based on collecting potential patterns found in a processed program. This allows evaluation of their frequency in a program and selection of patterns that may be interesting for a programmer. Basically, compared to the first method (by comparing), it is unique in its ability to recognize new patterns differing from each other in their subtrees, not only in their leaves. As opposed to the previous experiment, the input consists of abstract syntax trees and not derivation trees.

The main idea of the proposed algorithm is to avoid direct comparison of program fragments with each other. Instead, a structure of a fragment should be described using *structural schema*. This is a data structure that reflects structure of program fragment. It is obvious that a fragment of program can correspond to several structural schemas describing it with different levels of detail.

In our case structural schema is implemented as a modified abstract syntax tree containing variables. Variables can match different subtrees allowing covering program fragments that differ whole subexpressions. Structural schemas are actually potential patterns and therefore they have the same structure.

Fig. 4 depicts the algorithm where input of the algorithm is a sequence of abstract syntax trees representing expressions of a program or collection of programs. In the first step structural schemas are derived from each expression tree. This is achieved in two steps:

1) For each tree a set of its subtrees is generated. This allows finding patterns at any level of a program.
2) For each subtree a set of structural schemas is generated based on all the possible modifications of a particular tree T. By modification, we mean substitutions of the tree leaves or nodes by variables.

Then, an associative array is created of which the keys are structural schemas and values are lists of trees or subtrees covered by a particular structural schema.

Within the associative array it is possible to determine multiple occurrences of structural schemas. These schemas can be considered as patterns. Moreover, it is possible to reduce the associative array by those schemas that represent generalization of other schemas, without higher frequency.

This algorithm has been evaluated within an experiment based on a simple language of functions and expressions. To simplify development and make the relation between the internal structure and surface syntax more direct, S-expressions were used in the experiment.

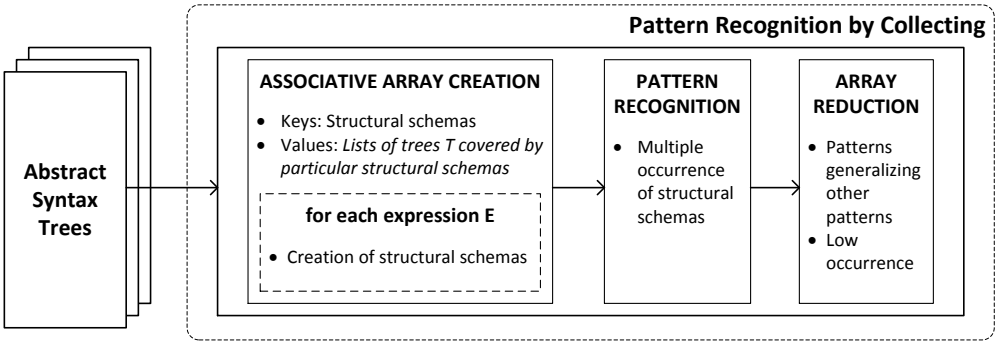


Fig. 4: Architecture of pattern recognition by collecting tools

For instance, if the following code sequence is used as an input for the algorithm:

```
(def squares xs
  (if (= xs nil) nil
            (cons (* (head xs) (head xs))
                  (squares (tail xs)))))
(def withTwo xs
  (if (= xs nil) nil
            (cons (+ (head xs) 2)
                  (withTwo (tail xs)))))
(def op xs
  (if (= xs nil) nil
            (cons (- (* (head xs) 2) 2)
                  (op (tail xs)))))
(def positives xs
  (if (= xs nil) nil
            (cons (>= (head xs) 0)
                  (positives (tail xs)))))
```

then it successfully recognizes the following pattern (variables marked with greek letters):

```
(def α xs (if (= xs nil) nil (cons β γ)))
```

V. Conclusion and Future Work

In this article, we have proposed a solution for automated introduction of new language abstractions based on patterns, which in this study are understood as recurring structures in program code. As part of the solution, two different approaches were experimentally developed to recognize language patterns: pattern recognition by comparing and by collecting.

The first approach is based on comparing program fragments alleging derivation trees of the applied Haskell grammar rules, generated by a complex set of analyzing tools [14]. Its principle lies in traversing particular derivation trees and recognizing the highest possible subtrees of the same structure.

While implementation of pattern recognition by comparing is relatively simple, and it is able to recognize most of the common recurring program structures, it does not allow substitution of particular subtrees by variables. That is, it cannot recognize some specific patterns. For instance, if we considered example code from the second experiment, the algorithm would only recognize the following pattern:

```
(def α xs (if (= xs nil) nil
    (cons (β (head xs) γ) (α (tail xs)))))
```

However, this covers only two of the four structurally similar fragments. More general pattern was supported and recognized only by the second pattern recognition technique, by collecting. It reduced drawbacks of the previous method, thus allowing substitution of subtrees by variables. Unlike in the previous experiment, the input of this algorithm consists of abstract syntax trees (not derivation trees) and its main principle lies in sequential generation of an associative array, with the keys of structural schemas and values of collected subtree lists corresponding to particular structural schemas.

Since it allows incremental addition of program expressions into the pattern recognition process, pattern recognition by collecting is also more suitable for interactive use as a part of programming environment. On the other hand, the need to generate extensive collection of structural schemas found in a program may have negative impact on the performance.

These experiments are significant to further part of the proposal focused on language enrichment using projection editing showing it is possible to find structural patterns in program code automatically. To make more significant conclusions, it is necessary to perform experiments on greater set of programs. Development of the language enrichment environment is needed to fully evaluate the proposal as well.

Further research in this area will be focused on the possibility to recognize patterns on higher level than the structure of code. These patterns may be scattered in the program code but semantically interconnected. Therefore, the pattern recognition process needs to have a high degree of knowledge about program semantics.

The contribution of the presented proposal for language enrichment is the new approach to the extension of programming language based on the needs of programmers [17]. It tries to combine the advantages of both linguistic and in-language abstractions, allowing language users to define new abstractions that are integrated into the language. In addition, the process of language enrichment is aided by automated patterns recognition.

However, upon the presented results, the most significant is the contribution to automated software evolution. Clearly, this would mean to shift from a language analysis to language abstraction, associating concepts to formal language constructs [18], and formalizing them by means of these associations. In this way, we expect to integrate programming and modeling, associating general purpose and domain-specific languages [19], [20], as well as to perform a qualitative move from an automatic roundtrip engineering [21], [22] to the automated roundtrip software evolution.

ACKNOWLEDGMENT

This work was supported by project VEGA 1/0341/13 Principles and methods of automated abstraction of computer languages and software development based on the semantic enrichment caused by communication.

REFERENCES

[1] F. P. Brooks, "No silver bullet: Essence and accidents of software engineering," *IEEE Computer*, vol. 20, no. 4, pp. 10–19, april 1987.

[2] M. Voelter, *DSL Engineering: Designing, Implementing and Using Domain-Specific Languages*. CreateSpace Publishing, 2013.

[3] M. P. Ward, "Language-oriented programming," *Software - Concepts and Tools*, vol. 15, no. 4, pp. 147–161, 1994.

[4] S. Dmitriev. (2004, November) Language oriented programming: The next programming paradigm. [Online]. Available: http://www.jetbrains.com/mps/docs/Language_Oriented_Programming.pdf

[5] H. Abelson and G. J. Sussman, *Structure and Interpretation of Computer Programs*, 2nd ed., ser. MIT Electrical Engineering and Computer Science. The MIT Press, 1996.

[6] D. Astapov, "Using haskell with the support of business-critical information systems," *Practice of Functional Programming*, vol. 2, 2009.

[7] A. Ott, "Using scheme in the development of "dozor-jet" family of products," *Practice of Functional Programming*, vol. 2, 2009.

[8] E. Gamma, R. Helm, R. Johnson, and J. Vlissides, *Design patterns: elements of reusable object-oriented software*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 1995.

[9] L. Kats, E. Visser, and G. Wachsmuth, "Pure and declarative syntax definition: Paradise lost and regained," in *Proceedings of Onward!* ACM, 2010.

[10] S. Erdweg, P. Giarrusso, and T. Rendel, "Language composition untangled," in *Proceedings of Workshop on Language Descriptions, Tools and Applications*, 2012.

[11] P. Graham, *On Lisp*. Prentice Hall, 1994. [Online]. Available: http://www.paulgraham.com/onlisp.html

[12] C. Simonyi, M. Christerson, and S. Clifford, "Intentional software," in *OOPSLA '06: Proceedings of the 21st annual ACM SIGPLAN conference on Object-oriented programming systems, languages, and applications*. New York, NY, USA: ACM, 2006, pp. 451–464.

[13] M. Fowler. (2005) Language workbenches: The killer-app for domain specific languages? [Online]. Available: http://martinfowler.com/articles/languageWorkbench.html

[14] E. Pietriková, L. Wassermann, S. Chodarev, and J. Kollár, "The effect of abstraction in programming languages," *Journal of Computer Science and Control Systems*, vol. 4, no. 1, pp. 137–142, 2011.

[15] J. Kollár, S. Chodarev, E. Pietriková, and L. Wassermann, "Identification of patterns through haskell programs analysis," in *Proceedings of the Federated Conference on Computer Science and Information Systems*. IEEE, 2011, pp. 891–894.

[16] J. Kollár, E. Pietriková, and S. Chodarev, "Abstraction in programming languages according to domain-specific patterns," *Acta Electrotechnica et Informatica*, vol. 12, no. 2, pp. 9–15, 2012.

[17] G. L. Steele, "Growing a language," *Higher-Order and Symbolic Computation*, vol. 12, pp. 221–236, 1999.

[18] J. Porubän and P. Václavík, "Extensible language independent source code refactoring," in *AEI '2008: International Conference on Applied Electrical Engineering and Informatics*, 2008, pp. 58–63.

[19] M. Sabo and J. Porubän, "Preserving design patterns using source code annotations," *Journal of Computer Science and Control Systems*, vol. 2, no. 1, pp. 53–56, 2009.

[20] I. Luković, P. Mogin, J. Pavićević, and S. Ristić, "An approach to developing complex database schemas using form types," *Software – Practice & Experience*, vol. 37, no. 15, pp. 1621–1656, Dec. 2007.

[21] U. Aßmann, "Automatic roundtrip engineering," *Electronic Notes in Theoretical Computer Science*, vol. 82, no. 5, pp. 33–41, 2003.

[22] C. Lohmann, J. Greenyer, and J. Jiang, "Applying triple graph grammars for pattern-based workflow model transformations," *Journal of Object Technology*, vol. 6, no. 9, pp. 253–273, 2007.

Modeling and Optimal Control of Nonlinear Underactuated Mechanical Systems – a Survey

[1]*Slávka JADLOVSKÁ (2nd year)*
Supervisor: [2]Ján SARNOVSKÝ

[1,2]Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

[1]slavka.jadlovska@tuke.sk, [2]jan.sarnovsky@tuke.sk

***Abstract*—The purpose of this paper is to examine the state-of-the-art in the analysis and control of nonlinear underactuated mechanical systems. Typical representatives of such systems are presented together with theoretical fundamentals. Application of optimal control techniques and hybrid systems theory to underactuated systems is described. Future research challenges are suggested and a list of essential references is included.**

***Keywords*—nonlinear underactuated mechanical systems, Lagrangian mechanics, optimal control techniques, hybrid systems theory**

I. Introduction

Underactuated systems represent a significant group of mechanical systems which range from simple planar robots or inverted pendulum systems to advanced higher-order systems with applications in robotics and air/sea transport [1]. In general, these systems are inherently nonlinear and have fewer control inputs than degrees of freedom [2], which presents a significant challenge to modeling and controller design [3].

This paper aims to provide a concise survey of the main achieved results and applications of underactuated systems with frequent references to crucial works in the field. After a brief summary of mathematical and physical preliminaries, principal examples of underactuated mechanical systems are presented. The ability of optimal control techniques to suit the properties of underactuated systems is next evaluated, and the potential of hybrid systems theory, which describes the integration of continuous/discrete dynamics in a dynamical system, is briefly examined with regard to modeling and control of underactuated systems.

All the way throughout the paper, open research problems are indicated. Most of these explore the theoretical and practical aspects of mutual overlaps between the physics of underactuated systems, optimal control techniques, and hybrid systems theory [1].

II. Modeling of Nonlinear Underactuated Systems using Lagrangian Mechanics

According to the Lagrangian formulation of classical mechanics, every possible configuration of a multi-body mechanical system can be described by a vector of generalized coordinates $\boldsymbol{\theta}(t)$, which correspond to the degrees of freedom (DoFs) of the system. Using the d'Alembert maximum principle, *Euler-Lagrange equations* were derived (one equation of motion is specified for every DoF) [1][4]:

$$\frac{d}{dt}\left(\frac{\partial L(t)}{\partial \dot{\boldsymbol{\theta}}(t)}\right)-\frac{\partial L(t)}{\partial \boldsymbol{\theta}(t)}+\frac{\partial D(t)}{\partial \dot{\boldsymbol{\theta}}(t)}=\boldsymbol{Q}^{*}(t) \quad (1)$$

where $L(t)$ is the difference between multi-body system's kinetic and potential energies (each given as a sum of energies of individual bodies), $D(t)$ stands for the dissipation properties and $\boldsymbol{Q}^{*}(t)$ is the vector of generalized external inputs. The process of mathematical model derivation via (1) is naturally transformable into a general algorithm which can be implemented using accessible symbolic software packages such as MATLAB's *Symbolic Math Toolbox*, *Maple*, and *Wolfram Mathematica* [1]. The mathematical model of a general controllable mechanical system derived from (1) is given as a following set of second-order differential equations:

$$\ddot{\boldsymbol{\theta}}(t)=\boldsymbol{f}\left(\boldsymbol{\theta}(t),\dot{\boldsymbol{\theta}}(t),\boldsymbol{u}(t),t\right) \quad (2)$$

The often-present assumption that the forward dynamics is affine in the direction of the produced torque yields a slightly constrained representation of the system:

$$\ddot{\boldsymbol{\theta}}(t)=\boldsymbol{f}_1\left(\boldsymbol{\theta}(t),\dot{\boldsymbol{\theta}}(t),t\right)+\boldsymbol{G}\left(\boldsymbol{\theta}(t),\dot{\boldsymbol{\theta}}(t),t\right)\boldsymbol{u}(t) \quad (3)$$

It is often useful to rearrange (3) into the standard *(minimal ODE – ordinary differential equation*) form [5]:

$$\boldsymbol{M}\left(\boldsymbol{\theta}(t)\right)\ddot{\boldsymbol{\theta}}(t)+\boldsymbol{N}\left(\boldsymbol{\theta}(t),\dot{\boldsymbol{\theta}}(t)\right)\dot{\boldsymbol{\theta}}(t)+\boldsymbol{R}\left(\boldsymbol{\theta}(t)\right)=\boldsymbol{V}(t)\boldsymbol{u}(t) \quad (4)$$

where $\boldsymbol{M}\left(\boldsymbol{\theta}(t)\right)$ is the inertia matrix, $\boldsymbol{N}\left(\boldsymbol{\theta}(t),\dot{\boldsymbol{\theta}}(t)\right)$ describes the influence of centrifugal /Coriolis forces, $\boldsymbol{R}\left(\boldsymbol{\theta}(t)\right)$ accounts for gravity forces and $\boldsymbol{V}(t)$ is the system input vector.

The system given as (3) or (4) is *fully actuated* in configuration $\left(\boldsymbol{\theta}(t),\dot{\boldsymbol{\theta}}(t),t\right)$ if it is able to command immediate acceleration in an arbitrary direction [3]:

$$rank\left(\boldsymbol{G}\left(\boldsymbol{\theta}(t),\dot{\boldsymbol{\theta}}(t),t\right)\right)=rank\left(\boldsymbol{V}\left(\boldsymbol{\theta}(t)\right)\right)=\dim\left(\boldsymbol{\theta}(t)\right) \quad (5)$$

If the range of directions in which immediate acceleration can be commanded is limited, the system is *underactuated*:

$$rank\left(\boldsymbol{G}\left(\boldsymbol{\theta}(t),\dot{\boldsymbol{\theta}}(t),t\right)\right)=rank\left(\boldsymbol{V}\left(\boldsymbol{\theta}(t)\right)\right)<\dim\left(\boldsymbol{\theta}(t)\right) \quad (6)$$

Typically, underactuated systems have fewer actuators than DoFs [1]. The difference $\dim\left(\boldsymbol{\theta}(t)\right)-rank\left(\boldsymbol{V}\left(\boldsymbol{\theta}(t)\right)\right)$ specifies the *degree of underactuation* of the system.

A. *Inverted Pendulum Systems*

Stabilization of a physical pendulum or a system of interconnected pendulum links in the upright unstable position is a benchmark problem in nonlinear control theory: in recent years, several types of stabilizing mechanisms such as cart moving on a rail [5], rotary arm [6], vertical oscillating base, or gyroscope have been introduced. Inverted pendulum systems (IPSs) are therefore regularly employed as typical examples of *unstable nonlinear underactuated systems* in the process of verification of linear or nonlinear control strategies in corresponding control structures [1][5][6]. Direct practical applications include walking robots, launching rockets, earthquake-struck buildings and two-wheel vehicles such as the *Segway PT* [1]. Principles of modeling and control of IPSs can further be considered as the basic starting point for the research of advanced underactuated systems such as mobile robots and manipulators [3], as well as aircraft / watercraft vehicles [2].

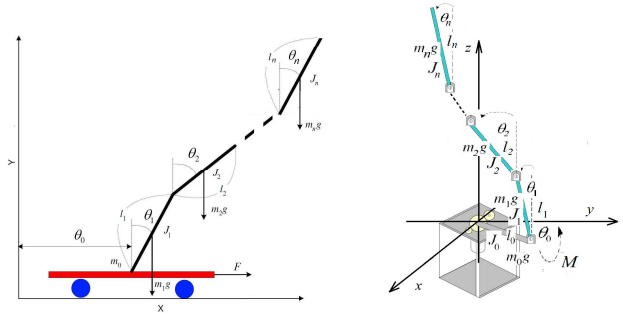


Fig. 1 Scheme and nomenclature: a) Generalized classical inverted pendulum system b) Generalized rotary inverted pendulum system

During our research, we focused on the mutual analogy among mathematical models of IPSs with various number of pendulum links. Consequently, we introduced the concept of a *generalized (n-link) inverted pendulum system* with n+1 DoFs and a single actuator, which allows to treat an arbitrary system of interconnected inverted pendula as a particular instance of the system of n pendula attached to a given stabilizing base (Fig. 1). A general procedure which determines the Euler-Lagrange equations of motion for a user-specified instance of a generalized classical and rotary IPS was developed and implemented via *Symbolic Math Toolbox* [7].

B. *Artificial Underactuated Systems*

Acrobot, *Pendubot* and the *inertia wheel pendulum* are all examples of underactuated systems with two degrees of freedom and a single actuator [2][8] which were introduced artificially to create complex low-order nonlinear dynamics and gain insight into control of high-order underactuated systems. Graphical representation of the *Acrobot* and the *Pendubot* is similar – both systems are depicted as two-link planar robots with revolute joints and share the same matrix of inertia. In the case of *Acrobot*, the actuator is placed at the elbow, while the *Pendubot* is actuated at the shoulder (Fig. 2).

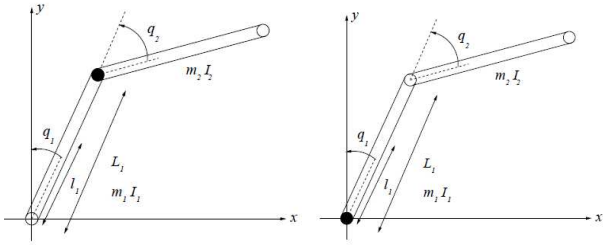


Fig. 2 Scheme and nomenclature: a) Acrobot b) Pendubot

The inertia wheel pendulum is composed of a physical pendulum with a rotating uniform inertia wheel at the end of the pendulum rod which is not directly actuated: in order to stabilize the pendulum in the upright equilibrium, the system has to be controlled via the rotating wheel.

C. *Mobile and Manipulator Robotic Systems*

Stabilization and tracking problems in *mobile robotics* generally involve underactuated mechanical systems [2][8]. If a robot with n inner connections and n actuators is not attached to the ground and instead performs walking, brachiating, gymnastic, swimming or flying motion [2], the number of its DoFs increases by the six DoFs which define its position and spatial orientation. Every additional control surface (i.e. a moveable platform) adds another actuator and a DoF to the system. *Robot manipulators* are often underactuated by construction, and a fully actuated manipulator becomes underactuated whenever the manipulated body provides the system with additional DoFs.

Principles of control for *underactuated systems* can also be employed to improve control of *fully actuated systems* either by increasing the effectivity of the employed actuators or by decreasing the design complexity [3]. After all, if the standard initial assumption of *rigid robotic arms* is omitted, we can claim that every robotic system is underactuated.

D. *Aircraft & Watercraft Systems*

Two significant groups of underactuated systems include aircraft (helicopters, airplanes, spaceships, satellites) and watercraft systems (ships, boats, submarines). Stabilization of the system in the direction of individual DoFs in the water/air environment, trajectory tracking and planning are specified as the principal analyzed problems. Underactuation is generally implied by the design and construction of a particular vehicle (Fig. 3). The *PVTOL* (*planar vertical take-off and landing*) airplane system is an underactuated system with three DoFs and two actuators which is often employed as a simplified planar model of the takeoff and landing of a helicopter [8]. The *helicopter/airplane system* is standardly described by six DoFs – position (x, y, z) and rotation angles along the three axes (*pitch* – lateral, *roll* – longitudinal, *yaw* – vertical rotation) and four control inputs – three control moments in the body frame and the main rotor thrust [2]. Ocean vessels are equipped with propellers and rudders which enable control in two directions only (*surge* – longitudinal, *heave* – vertical axis) without any direct control in the direction of lateral motion (*sway*) [9]. The degree of underactuation can also increase in the case of actuator failure, or if the number of motors is intentionally reduced to decrease the load mass onboard the aircraft or watercraft.

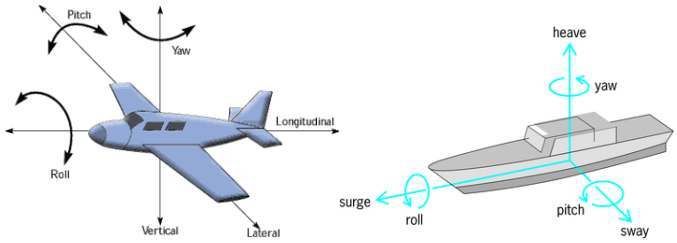


Fig. 3 Scheme and degrees of freedom related to principal axes: a) Airplane b) Ship

Living systems are often underactuated in the interaction with their surroundings. Although the human body contains more *actuators* (muscles) than *DoFs* (joints), it can be easily proven that our body is underactuated despite the presence of fully actuated joints: if we jump into air, no combination of control inputs from muscles is able to counter the influence of gravity and aerodynamic forces and change the trajectory of our center of gravity [3]. Similar considerations apply to the other living creatures. A flying bird, a swimming fish and a walking human are all examples of mechanical systems whose locomotion is due to changes in their physical shape, leading to indirect position actuation. As it is shown in [10], by studying various types of animal locomotion and the way they overcome their own underactuated dynamics, we can get significant inspiration for design of underactuated vehicles.

III. Optimal Control of Nonlinear Underactuated Systems

Fully actuated systems possess a number of strong structural properties which facilitate design of optimal/robust/ adaptive controllers, e.g. feedback linearizability, passivity and linear parametrizability. These are usually lost in underactuated systems, while at the same time undesirable properties (higher relative degree, nonminimum phase behavior) emerge [2]. Control of underactuated systems subsequently becomes more difficult, with fewer general results available.

It has been shown that optimal control techniques yield reliable, consistent results for underactuated systems. The goal of optimal control design for a *linear, time-invariant* dynamic system is to determine such feedback control so that a given criterion of optimality is achieved [11]. In case the considered linear system is actually a linear approximation of a nonlinear system around a given equilibrium, then the optimal techniques for linear systems yield an approximate, locally near-optimal stabilizing solution with guaranteed closed-loop stability and robustness. In [1][12], we solved the problem of IPSs stabilization in the unstable position via optimal control algorithms based on quadratic functional minimization, using continuous-time and discrete-time linearized state-space models of IPSs. We also introduced additional control structures which ensure that the cart/arm position reaches the reference value (by means of feedforward gain), and the permanent steady-state error is eliminated (by implemented integral control).

Model predictive control (MPC) is a discrete-time optimal control technique in which the control action for each time step is computed by solving an on-line optimization problem in finite time (*receding horizon control*) while at the same time considering input/state constraints [13]. MPC is currently the only modern control technique with significant impact on industrial process control: compared to the 1980s, when MPC technology first became popular in petrochemical industry, commercial MPC implementations can now be found in chemical and food processing, automotive and aerospace applications [15]. However, application of the MPC algorithm for most nonlinear underactuated systems still presents a challenge in terms of disturbance/steady-state error elimination. To solve problems arising from the structure of an underactuated system, suitable adjustment of the MPC algorithm [13][14] is required.

IV. Hybrid Systems Theory in Modeling and Control of Underactuated Systems

To provide a convenient framework for modeling and control of systems characterized by an interaction between continuous (*time-driven*) and discrete (*event-driven*) dynamics, *hybrid systems theory* was developed [16]. As a result, various engineering problems which were once considered a case of a particular implementation can now be researched systematically as part of a complex theory.

A. Modeling of Underactuated Hybrid Systems

Hybrid models are often useful if we have to consider discontinuous development of the mechanical system dynamics, i.e. a robotic arm whose continuous motion is interrupted by collisions or strikes to the surface, or if the arm dynamics is subject to state jumps caused by the arm shooting out objects. In [17], a hybrid model of such an underactuated robot is defined as a modification of the standard minimal form (which includes an operator describing the jumps in the state vector), and employed in a trajectory planning algorithm. Out of the available mathematical formalisms (modeling frameworks) of hybrid systems, *PWA* and *MLD* forms are often employed: the *PWA* form interconnects the linear state-space representation and discrete automata, dividing the input/state space into regions defined by polyhedra, and the *MLD* form is composed of a system of linear difference equations which can assume real and binary values and a set of linear inequalities to describe the constraints [18].

B. Control of Underactuated Hybrid Systems

Optimal control theory is the principal approach to hybrid systems control, and the complexity of an optimal control problem decreases if the system is expressed in discrete-time, since its main source in a hybrid system is the number of possible switching scenarios. Optimal control problems can be solved for hybrid systems in the discrete-time state-space form using either *PWA* or *MLD* models, which was first outlined by Sontag in [19]. It is next shown by Borrelli [18] that the solution of the optimal control problem in finite time is a time-variant piecewise-affine feedback control law, defined over non-convex regions. *Hybrid model predictive control*, which has recently attracted much attention, is suitable for systems defined by switched linear dynamics which are subject to linear/logical constraints on state/input variables [20].

Application of hybrid optimal/predictive control algorithms on underactuated systems has already been covered by multiple authors, although no consistent approach has yet been developed. In a survey paper, Buse et al. [21] demonstrate the application of hybrid optimal control techniques in control of an underactuated robot arm. Yin & Hosoe [22] employ hybrid predictive control to plan the trajectory of a walking robot expressed in *MLD* form, and Rodrigues & How [23] develop an algorithm to automate the transformation of IPSs into the *PWA* representation to enable subsequent hybrid control.

The advancement of hybrid systems theory is supported by a wide range of available software tools which enable symbolic/numeric computations and simulations in accordance with theoretical results. Most of these have been developed by research groups led by *Prof. Morari* of ETH Zürich and *Prof. Bemporad* of ETH Zürich, later University of Siena [18][20], and include tools which simplify the process of formulation

and analysis of a hybrid model (*HYSDEL* modeling language), enable experimental identification of hybrid models (*Hybrid Identification Toolbox* (HIT)), and provide functions for hybrid optimal/predictive control algorithm design as well as closed-loop simulation (*Hybrid Toolbox* (HT) [24], and *Multiparametric Toolbox* (MPT) [25]).

From a different viewpoint, Liberzon [26] presents examples of control problems of underactuated systems where it is necessary or useful to employ switching control structures. If the desired trajectory of the underactuated system consists of multiple pieces of significantly different parts (e.g. aircraft maneuvers) or if the state space contains obstacles, we might need to choose different controllers at different stages of the problem and implement switching between them. Also, switching control is often the only feasible way to control a nonholonomic system, since there is no continuous control which could stabilize such systems on a given time interval.

V. Conclusion

This paper presents a compact summary of results which have so far been achieved in modeling and optimal control of nonlinear mechanical underactuated systems using classical and hybrid approaches. Great practical importance of underactuated systems in mobile robotics, aviation and ship transport has sparked much interest from physicists and control theorists alike. The principal aim of currently conducted research is to overcome the difficulty of control algorithm design caused by certain disadvantageous physical properties of underactuated systems.

After a brief survey on fundamental principles of mechanical system modeling based on Lagrangian mechanics, an overview of principal categories of underactuated systems was presented. For each category, the reason for underactuation was specified together with control objectives addressed in referenced works. Optimal control was confirmed as a reliable control technique for underactuated systems; it was shown that adjustments to predictive algorithms are required. Application areas of hybrid systems theory were discovered to include hybrid models for underactuated systems with logical parts, as well as hybrid optimal/predictive control algorithms and switching control structures. Numerous problems with future research potential were emphasized in the paper.

The findings presented in this paper are elaborated in the referenced thesis for dissertation examination which describes the proposed integration of theories of underactuated systems, optimal control, and hybrid systems. As a meaningful contribution to modeling/control education, nonlinear underactuated systems are being integrated into the research and teaching activities of the Center of Modern Control Techniques and Industrial Informatics at the DCAI-FEEI TU.

Acknowledgment

This contribution is the result of the Vega project implementation – Dynamic Hybrid Architectures of the Multi-agent Network Control Systems (No. 1/0286/11), supported by the Scientific Grant Agency of Slovak Republic – 70%, and of the KEGA project implementation – CyberLabTrainSystem – Demonstrator and Trainer of Information-Control Systems (No. 021TUKE-4/2012) – 30%.

References

[1] S. Jadlovská, Modeling and Optimal Control of Nonlinear Underactuated Dynamical Systems [Modelovanie a optimálne riadenie nelineárnych podaktuovaných dynamických systémov]. Thesis for Dissertation Examination. Košice: FEEI-TU, 2013.

[2] M. W. Spong, "Underactuated Mechanical Systems: Control Problems in Robotics and Automation", in *Lecture Notes in Control and Information Sciences*, vol. 230, 1998, pp. 135-150.

[3] R. Tedrake, Underactuated Robotics: Learning, Planning and Control for Efficient and Agile Machines. Course Notes for MIT 6.8., Cambridge: Masachusetts Institute of Technology (MIT), 2009.

[4] H. Goldstein, Ch. Poole, J. Safko, *Classical Mechanics*, 3rd ed. Addison-Wesley, 2001. 680 p., ISBN 978-0201657029.

[5] A. Bogdanov, Optimal Control of a Double Inverted Pendulum on the Cart. Technical Report CSE-04-006, OGI School of Science and Engineering, OHSU, 2004.

[6] K. Furuta, M. Yamakita, S. Kobayashi, "Swing Up Control of Inverted Pendulum", *Proc. of the Int. Conf. on Industrial Electronics, Control and Instrumentation (IECON'91)*, Kobe, Japan, Oct 28-Nov 1, 1991.

[7] S. Jadlovská, J. Sarnovský, "Modelling of Classical and Rotary Inverted Pendulum Systems – a Generalized Approach", in *Journal of Electrical Engineering*, vol. 64, no. 1, 2013, pp. 12–19, ISSN 1335-3632.

[8] R. Olfati-Saber, Nonlinear Control of Underactuated Mechanical Systems with Application to Robotics and Aerospace Vehicles. PhD Thesis. Cambridge: Massachusetts Institute of Technology (MIT), Department of Electrical Engineering and Computer Science, 2001.

[9] K. D. Do, "Practical Control of Underactuated Ships", in *Ocean Engineering*, vol. 37, 2010, pp. 1111-1119.

[10] W. Li, E. Todorov, "Iterative Linear-Quadratic Regulator Design for Nonlinear Biological Movement Systems", *Proc. of the 1st International Conference on Informatics in Control, Automation and Robotics*, vol. 1, 2004, pp. 222-229.

[11] F. L. Lewis, D. L. Vrabie, V. L. Syrmos, *Optimal Control*. Wiley, 2012. 552 p, ISBN 978-0470633496.

[12] S. Jadlovská, J. Sarnovský, A Complex Overview of Modeling and Control of the Rotary Single Inverted Pendulum System, Advances in Electrical and Electronic Engineering, 2013, ISSN 1336-1376, in press.

[13] G. G. Goodwin, M. M. Seron, J. A. De Doná, *Constrained Control and Estimation : An Optimisation Approach*, London: Springer, 2005. 429 p., ISBN 978-1849968836.

[14] A. C. D. Caldeira, F. A. C. C. Fontes, "Model Predictive Control of Under-actuated Mechanical Systems", *Proc. of ROBOMAT'07*, September 17-19, 2007, Coimbra, Portugal.

[15] J. S. Qin, T. A. Badgwell, "A Survey of Industrial Model Predictive Control Technology", in *Control Engineering Practice*, vol. 11, 2003

[16] J. Lunze, F. Lamnabhi-Lagarrigue, *Handbook of Hybrid Systems Control: Theory, Tools, Applications*. New York: Cambridge University Press, 2009. 582 p., ISBN 978-0521765053.

[17] P. X. La Hera, A. S. Shiriaev, L. B. Freidovich, U. Mettin, S. V. Gusev, "Stable Walking Gaits for a Three-Link Planar Biped Robot with One Actuator", in *IEEE Trans. on Robotics*, in press.

[18] F. Borrelli, A. Bemporad, M. Morari, *Predictive Control For Linear and Hybrid Systems*. Berkeley: University of California, 2012.

[19] E. D. Sontag, "Nonlinear Regulation: The Piecewise Linear Approach", in *IEEE Trans. on Automatic Control*, vol. 26, no. 2, April 1981.

[20] A. Bemporad, "Control of Systems Integrating Logic, Dynamics and Constraints", in *Automatica*, vol. 35, no. 3, 1999, pp. 402-427, ISSN 0005-1098.

[21] M. Buse, M. Glocker, M. Hardt, O. Von Stryk, R. Bulirsch, G. Schmidt, "Nonlinear Hybrid Dynamical Systems: Modeling, Optimal Control and Applications", in *Lecture Notes in Control and Information Sciences*, vol. 279, Berlin, Heidelberg: Springer-Verlag, 2002, pp. 311-335.

[22] Y. Yin, Sh. Hosoe, "A Hybrid System Control Approach to Biped Robot Control", in *Journal of Automation, Mobile Robotics and Intelligent Systems*, vol. 2, no. 4, 2008.

[23] L. Rodrigues, J. P. How, "Automated Control Design for a Piecewise-Affine Approximation of a Class of Nonlinear Systems", *Proc. of the Americal Control Conference*, Arlington, VA, June 25-27, 2001.

[24] A. Bemporad, Hybrid Toolbox for Real-Time Applications: User's Guide. Siena: University of Siena, Department of Information Engineering and Mathematical Sciences.

[25] M. Kvasnica, P. Grieder, M. Baotić, F. J. Christophersen, Multi-parametric Toolbox (MPT), Zürich: Institut für Automatik, ETH – Swiss Federal Institute of Technology, 2006

[26] D. Liberzon, *Switching in Systems and Control*. Birkhäuser, 2003. 248 p., ISBN 978-0817642976.

Modification of IDEFØ box model for the purpose of experimental identification effectiveness evaluation

[1]*Vladimír GAŠPAR (2nd year)*
Supervisor: [2]Ladislav MADARÁSZ

[1,2] Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

[1]vladimir.gaspar@tuke.sk, [2]ladislav.madarasz@tuke.sk

***Abstract*—This paper presents the possibility of functional model modifications for experimental identification (EI) process description, efficiency evaluation and visualization purpose. It describes basics of IDEFØ box conceptual modeling approach, states its benefits and drawbacks. A comparison between IDEFØ and IDEF3 models and their suitability for specific purpose of EI description is also denoted. General topic of the paper is the proposition of a functional and relation chaining modification of IDEFØ conceptual model. Using proposed approach, we are able to create a complex EI conceptual process model from both functional and time related point of view. This modification also enables researchers to predict the continuity and discover new phases of the identification. Moreover, when plotting model boxes onto multiple axes, the conceptual model may represent the EI process using time variable, other dependent variables, up to three dimensions for visualization and "n" dimensions only for evaluation (analytical) purposes.**

***Keywords*— efficiency, experimental identification description, IDEFØ, conceptual model modification.**

I. Introduction

In the past, EI was usually accompanied by detailed analysis of each of identification phases and the creation of final experimental model was considered the only aim. Unsuccessful identification phase held the place of main responsibility for low final model accuracy.

Nowadays, EI of complex systems also considers the details of identification process, to be responsible for the quality and accuracy of final experimental model [10][12]. This means that except for quantitative parameters (time, expenses), qualitative parameters (system efficiency, reliability and knowledge inside the EI process) also affect the final experimental model accuracy in a great extent. Existence of a relation between various parameters and model accuracy was denoted by Isermann [2] (see Fig. 1). Each part of the figure shows the relation between model accuracy and specific required inputs (knowledge or expenses). Fig. 1 also shows the difference between the character of the system (T.S. - technical system, M. a E.S. - mechanical and electronic system, EN. S. - energetic system). However, this explanation only considers the extent of the knowledge of system's internal processes, disregarding other qualitative parameters that have been mentioned before. References [2][13] focused only on the comparison between analytical and experimental identification and its differences, according to specific parameters. Their goal was not the evaluation and measurement of qualitative parameters but rather the identification process itself. These are the main reasons, why it is essential to evaluate qualitative parameters and to denote the structure of the EI process, which is considered the main novelty contributed by this paper. The role of IDEFØ in the whole evaluation process is described in the chapter III.

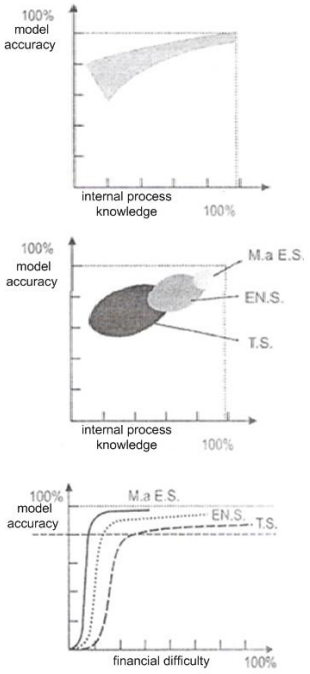


Fig. 1. Behavior of experimental model accuracy of a specific system, according to two parameters

II. IDEFØ and IDEF3 models comparison and the possibility of usage in EI efficiency evaluation

For the purpose of modeling, each one of "Integrated Computer Aided Manufacturing definition" models is suitable for different problems. According to the following simple study, we have chosen IDEFØ as the most flexible solution for modifications.

A. IDEFØ

IDEFØ is a definition for functional modeling of systems using box model representation. It identifies basic attributes that have an effect on the described system. It also enables creating simple relations between boxes. IDEFØ is suited for hierarchical organization of system elements and only involves

functional relations with no timestamps included [12][14]. It is usually used either to specify requirements of the system and also its specifications, if used as a preliminary modeling approach or it can also be used as an analyzing tool to describe an existing system. The combination of both approaches is possible and is also used as a start-point in case of EI modeling.

B. IDEF3

IDEF3 is considered a complementary modeling approach to IDEFØ [12][14]. Its aim is to describe state transitions and process flow according to predefined scenario. The application of this approach is mainly oriented on business process modeling rather that systems modeling. Unlike IDEFØ the aim is to describe the scenario of a process not each element and its behavior. This is why the application of IDEF3 modeling approach would not meet the need for EI phase description.

III. ROLE OF IDEFØ IN EXPERIMENTAL IDENTIFICATION PROCESS DESCRIPTION

Evaluation of a qualitative parameter and its transformation into a quantitative parameter always involves a particular amount of subjectivity and is usually estimated using other measurable parameters. The most common approach is the experimentation, where each test of an experiment is evaluated in binary as successful (denoted by sign "1") or unsuccessful (denoted by "0"), according to planned outputs [6][7][8][9][10]. This approach can only be used when there is a clear expectation of results that need to be obtained, for the test to be considered as successful.

However, usage of this method also requires to possess knowledge about parameters that influence particular identification phase and also the genesis of the identification before the actual phase [8]. To discover and describe these parameters, an IDEFØ box model can be beneficial and used as a source of various information in the future research [8]. The basic IDEFØ model uses inputs(I), outputs(O), constraints (C-control), mechanisms(M) to describe the element (see. Fig. 2)[12][14]. This means that the conceptual modeling of the identification process using IDEFØ is a necessary prerequisite for qualitative parameters evaluation and the identification process itself.

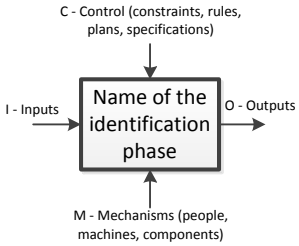


Fig. 2. General IDEFØ box model depiction

Using this approach, we described a phase of identification of the MPM 20 turbojet engine, as an example (see. Fig. 3).

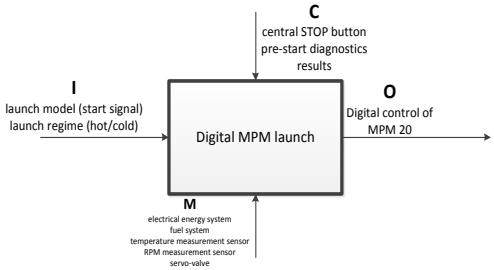


Fig. 3. Digital MPM launch identification phase, described in IDEFØ

There is also a possibility of searching for more detailed description. This description of the process takes much more time to create but is more beneficial, mainly when searching for a problem element in the system. An example depiction of detailed phase of MPM 20 EI model can be seen in Fig. 4 [8].

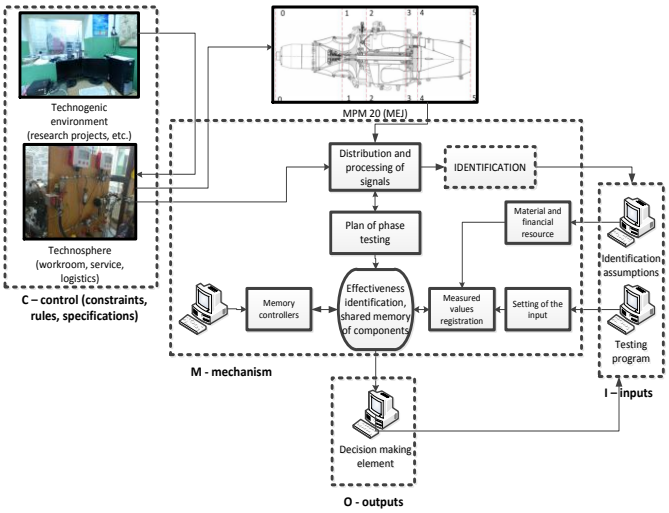


Fig. 4. Detailed description of an identification phase according to IDEFØ

If we describe every phase in this manner and find relations between phases, we may create a complex model including relations between each couple of phases. These relations may vary, depending on the character of the identification phase. If we consider the need for continuous efficiency increase, a feedback can also be added to the model. This feedback allows to stop the identification process, if the efficiency (output value) is decreasing between phases. It is considered as the mechanism of the previous phase (aposterior information) and actual (or future) phase, when using efficiency estimation. The continuity of the identification process can be predicted using aprior and aposterior information [8].

Moreover, the identification of I,C,O,M helps to find the cause of decreasing efficiency. An example depiction of box model chaining using relations with feedback, aprior and aposterior information is presented in Fig. 5.

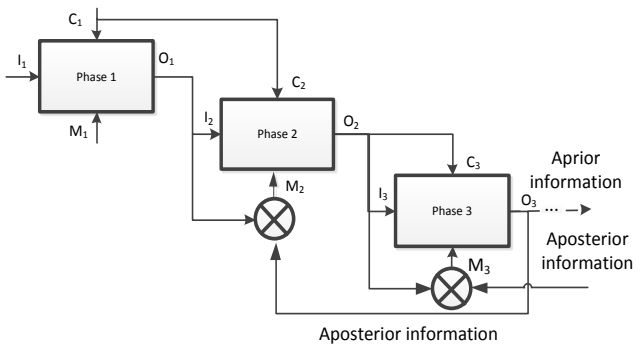


Fig. 5. Chaining of IDEFØ box models with relations and feedback

This conceptual model is sufficient for the purpose of experimental identification process description. However, in some cases, more than efficiency comparison is needed. In such a case, we may also consider other modifications.

Plotting Fig. 5 into a Cartesian coordinate system can add even more output information of this conceptual model. The time can be denoted on the x-axis as usual, and a dependent variable on "y" or "z" axes (in case of a 3D model).

The addition of x-axis makes the width parameter of each box useful. It denotes the time length of the specific phase. An example of the axis addition is shown in Fig. 6.

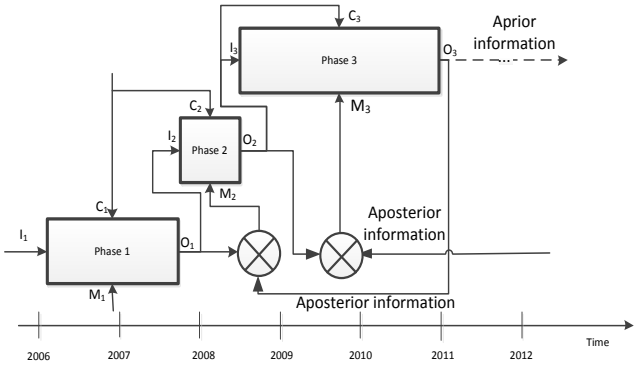


Fig. 6. Chaining of IDEFØ box models with relations, feedback and time parameter on the x-axis

Although, the addition of each axis brings more complexity to the conceptual model, it also creates a higher information value. For the purpose of this modification, transformed universal qualitative parameters are suitable to be plotted as dependent variables. These parameters may be evaluated using various methods, including the binary rating method, which has been briefly described in the beginning of the chapter III.

When using "y" and "z" axes, it is also possible to use box models (same as x-axis only, see Fig. 6). By each dimension value of the box, we can denote the interval of particular dependent variable (same approach as for the time variable, see Fig. 6) that describes the minimum and maximum value of the observed parameter (safety, reliability, efficiency, etc.), during the whole phase. Other possible solution is to use central point of the box, or to create a single point instead. This notation can be used, if we wish only to denote a single (mean, central, maximal, minimal) value of the respective dependent parameter. Again we consider important to note that this type of conceptual model does not describe the system as an experimental model but only enables better overview on the experimental identification process, relations between its phases, its efficiency progress, etc.

Adding more than two dependent parameters is possible but only for analytical purposes. To ensure the presence of the main advantage (design) only the maximum of 2 dependent and one independent variables can be plotted.

IV. Conclusion

Experimental identification of complex systems and its efficiency is a wide problem area. Moreover, in complex system it is difficult to understand the whole process of the identification and the relations between its phases, until it is described in details. In this paper we presented an approach using basics of functional and process conceptual modeling according to IDEFØ model. We also described disadvantages and advantages of IDEF3 model and its possibilities of usage for EI description. Modifications that have been carried out enable researchers not only to depict the identification process, relations and involved parameters but also allow to monitor the EI process, according to its efficiency. The last chapter of this paper also presents the possibility to add axes to the conceptual model solution and depict other transformed qualitative parameters in an easy and understanding way. All examples that have been shown in this paper were created for the purpose of experiments and EI in conditions of the Laboratory of intelligent control systems of jet engines.

Acknowledgment

The work presented in this paper was supported by VEGA, Grant Agency of Ministry of Education and Academy of Science of Slovak Republic under Grant No. 1/0298/12 – "*Digital control of complex systems with two degrees of freedom*". The work presented in this paper was also supported by KEGA under Grant No. 018TUKE-4/2012 – "*Progressive methods of education in the area of control and modeling of complex systems object oriented on aircraft turbo-compressor engines*.

References

[1] ČERNEJ, M., MADARÁSZ, L., GAŠPAR, V.: Economic aspects of identification of a small turbojet engine – 1 electronic optical disc (CD-ROM). In: Electrical Engineering and Informatics III : Proceeding of Faculty of Electrical Engineering and Informatics of the Technical University of Košice : September, 2012 Košice, Slovakia. pp. 52-58. – ISBN 978-80-553-0890-6.

[2] ISERMANN, R.: Identifikation Dynamischer Systeme 1: Grundlegende Methoden. Springer - Verlag, Berlin, Heidelberg. 1992. 330pp. ISBN 9780387549248.

[3] KELEMEN, M., LAZAR, T., KLECUN, R.: Ergatické systémy a bezpečnosť v letectve. Edukácia a inteligencia zručností v leteckej prevádzke. (Ergatic systems and safety in aviation. Education and intelligence of skills in aviation operation). Akadémia ozbrojených síl generála M. R. Štefánika, Liptovský Mikuláš. 2009. 316pp. ISBN 978-80-8040-383-6.

[4] KRAJŇÁK, P.: Využitie záznamu parametrov letu na hodnotenie technického stavu leteckého turbokompresorového motora. (Usage of flight parameter logging for jet engine technical state evaluation). Doktorandská dizertačná práca, KA LF TU Košice. 2010. 106 pp.

[5] LAZAR, T., BRÉDA, R., KURDEL, P.: Inštrumenty istenia letovej bezpečnosti (Instruments of flight security ensurement). Popradská tlačiareň, vydavateľstvo s.r.o. Košice 2011. 232pp. ISBN 978-80-553-0655-1.

[6] LAZAR, T., et al.: Tendencie vývoja a modelovanie avionických systémov. MO SR Bratislava, 2000, ISBN 80-88842-26-3, 160 pp.

[7] LAZAR, T., MADARÁSZ, L., ANDOGA, R., GAŠPAR, V.: Experimental Identification of a Small Turbojet Engine– 2012. 8pp. In:

proceedings II. Vedecká konferencia doktorandov, Faculty of Aeronautics, TU, Košice, Slovak Republic. ISBN 978-80-553-0914-9.

[8] LAZAR, T., MADARÁSZ, L., GAŠPAR, V.: Estimation process analysis of identification efficiency of small turbojet engine with intelligent control, elfa, s.r.o., Košice. 2013. 160pp. ISBN 978-80-8086-200-8 .

[9] LAZAR, T., MADARÁSZ, L., GAŠPAR, V.: The Efficiency of Experimental Identification of Cognitive Systems – 2013. In: proceedings SAMI 2013 Herľany, Slovak Republic. ISBN 978-1-4673-5927-6.

[10] LAZAR, T., MADARÁZ, L. et al.: Inovatívne výstupy z transformovaného experimentálneho pracoviska s malým prúdovým motorom (Innovative outputs from the transformed experimental laboratory with a small turbojet engine) elfa, s.r.o., Košice. 2011. 348 pp. ISBN 978-80-8086-170-4.

[11] MADARÁSZ, L.: Inteligentné technológie a ich aplikácie v zložitých systémoch (Intelligent Technologies and their applications in complex systems), University Press elfa, 2004. 348 pp. ISBN 80-89066-75-5.

[12] Manual - ICAM architecture Part II-Volume IV - Function Modeling Manual (IDEFØ), AFWAL-TR-81-4023, Materials Laboratory, Air Force Wright Aeronautical Laboratories, Air Force Systems Command, Wright-Patterson Air Force Base, Ohio 45433, June 1981.

[13] NOSKIEVIČ, P.: Modelování a identifikace systémú. MONATEX a.s. 1999, 154pp.

[14] Standard - Integration definition for function modeling (IDEFØ), Federal Information Processing Standards Publication (FIPS) 183, 21.12.1993.

The crossing number of the Cartesian product of the special graph on six vertices with the path P_n

Jana PETRILLOVÁ (4^{th} year)
Supervisor: Marián KLEŠČ

Dept. of Mathematics and Theoretical Informatics, FEI TU of Košice, Slovak Republic

jana.petrillova@tuke.sk

***Abstract*—The crossing numbers of Cartesian products $G\Box P_n$ for the path P_n of length n for all connected graphs G on at most five vertices are known. Moreover, the crossing numbers of Cartesian products of cycles or stars with all graphs of order at most four are also known. In this paper, we extend these results by determining the crossing number of the Cartesian product $G\Box P_n$ for the special graph G of order six.**

***Keywords*—Cartesian product, crossing number, drawing, graph**

I. Introduction

Let G be a simple graph with vertex set $V(G)$ and edge set $E(G)$. A drawing of a graph is a mapping of a graph into a surface. The crossing number $\mathrm{cr}(G)$ of a simple graph G is defined as the minimum possible number of edge crossings in a drawing of G in the plane. A drawing with minimum number of crossings (an optimal drawing) must be a good drawing; that is, each two edges have at most one point in common, which is either a common end-vertex or a crossing. Moreover, no three edges cross in a point. Let D be a good drawing of the graph G. We denote the number of crossings in D by $\mathrm{cr}_D(G)$. Let G_i and G_j be edge-disjoint subgraphs of G. We denote by $\mathrm{cr}_D(G_i, G_j)$ the number of crossings between edges of G_i and edges of G_j. In a good drawing D of the graph G, we say that a cycle C separates the cycles C' and C'' (the vertices of a subgraph G not containing vertices of C) if C' and C'' (the vertices of G) are contained in different components of $\mathbb{R}^2 \setminus C$. A region in nonplane drawing of any graph is taken in such a way that crossings are considered to be vertices in the plane. The Cartesian product $G_1\Box G_2$ of graphs G_1 and G_2 has vertex set $V(G_1\Box G_2) = V(G_1) \times V(G_2)$ and any two vertices (u, u') and (v, v') are adjacent in $G_1\Box G_2$ if and only if either $u = v$ and u' is adjacent with v' in G_2, or $u' = v'$ and u is adjacent with v in G_1.

The investigation on the crossing number of graphs is a classical and very difficult problem. The crossing numbers has been studied to improve the readability of hierarchical structures. A crossing of two edges of the communication graph requires unit area in VLSI-layout. So, the crossing number together with the number of vertices of the graph immediately provide a lower bound for the area of the VLSI-layout of the communication graph. For that reason the problem of crossing numbers was studied also by VLSI communities and computer scientists.

According to their special structure, Cartesian products of special graphs are one of few graph classes for which the exact values of crossing numbers were obtained. Let C_n be the cycle of length n, P_n be the path of length n, and S_n be the star isomorphic to $K_{1,n}$. Beineke and Ringeisen [1] started to study the crossing numbers of Cartesian products of cycles with all graphs on at most four vertices. The crossing numbers of Cartesian products of cycles, paths and stars with all 4-vertex graphs are determined in [3], [4], and [5]. The crossing numbers of Cartesian products of paths with all graphs of order five are collected in [8]. It seems natural to enquire about crossing numbers of Cartesian products of paths with other graphs. There are known the crossing numbers of products $G\Box P_n$ for some 6-vertex graphs G, see [6], [10], [11], [12], [13], and [14]. The crossing numbers of Cartesian products of paths with 40 graphs of order six are collected in [9]. Let G be the connected graph of order six which consists of one 4-cycle and one path of length 2. In this paper, we extend the results in [9] by determining the crossing number of the Cartesian product $G\Box P_n$.

II. The main result

We assume $n \geq 1$ and find it convenient to consider the graph $G\Box P_n$ in the following way: it has $6(n+1)$ vertices and edges that the edges are in $n+1$ copies $G^{(i)}$, $i = 0, 1, \ldots, n$, and in six paths of length n. For $i = 0, 1, \ldots, n$, let a_i, b_i, c_i and e_i be the vertices of $G^{(i)}$ of degree two, d_i the vertex of degree three and f_i the vertex of degree one (see Fig. 1). Thus, for $x \in \{a, b, c, d, e, f\}$, the path $P_n^{(x)}$ is induced by the vertices $x_0, x_1, \ldots, x_n$. For $i = 0, 1, 2, \ldots, n-1$, let $M^{(i)}$ denote the subgraph of $G\Box P_n$ containing the vertices of $G^{(i)}$ and $G^{(i+1)}$ and six edges joining $G^{(i)}$ to $G^{(i+1)}$. Let $Q^{(i)}$, $i = 1, 2, \ldots, n-1$, denote the subgraph of $G\Box P_n$ induced by $V(G^{(i-1)}) \cup V(G^{(i)}) \cup V(G^{(i+1)})$. So, $Q^{(i)} = G^{(i-1)} \cup M^{(i-1)} \cup G^{(i)} \cup M^{(i)} \cup G^{(i+1)}$. Let us denote by $C_4^{(i)}$, $i = 0, 1, \ldots, n$, the subgraph of $G^{(i)}$ obtained from $G^{(i)}$ by removing two vertices e_i and f_i. If the edges of some 4-cycle of the graph $G\Box P_n$ cross each other, we say that this 4-cycle has an internal crossing. The edges of a 4-cycle cannot cross more than once in a good drawing.

The graph $G\Box P_1$ is planar. The crossing number of the graph $G\Box P_2$ is one, because the graph $S_3\Box P_2$ is its subgraph and $\mathrm{cr}(S_3\Box P_2) = 1$ (see [3]). The reverse inequality $\mathrm{cr}(G\Box P_2) \leq 1$ follows from the drawing in Fig. 2. The next lemma help us to prove that the crossing number of the graph $G\Box P_n$ is $2(n-2)$ for $n \geq 3$.

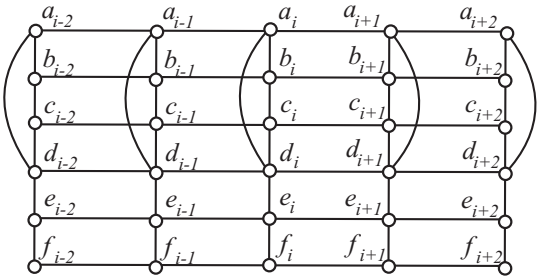


Fig. 1. The graph $G^{(i-2)} \cup M^{(i-2)} \cup Q^{(i)} \cup M^{(i+1)} \cup G^{(i+2)}$.

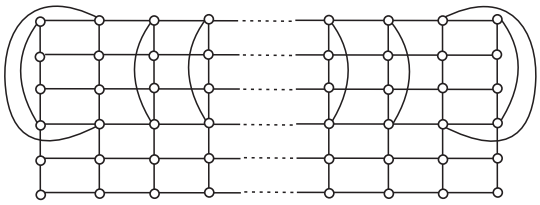


Fig. 2. The drawing of the graph $G\square P_n$ with $2(n-2)$ crossings.

Lemma 1 *If D is a good drawing of the graph $G\square P_n$, $n \geq 3$, in which every of the subgraphs $G^{(i)}$, $i = 0, 1, 2, ..., n$, has at most one crossing on its edges, then in D there are at least $2(n-2)$ crossings.*

Proof. The aim of this proof is to show that if every of the subgraphs $G^{(i)}$, $i = 0, 1, 2, ..., n$, has at most one crossing on its edges, then every subgraph $Q^{(i)}$, for all $i = 2, 3, ..., n-2$, has at least two crossings and each $Q^{(1)}$ and $Q^{(n-1)}$ has at least one crossing, which are not counted for other subgraph $Q^{(j)}$, $i \neq j$.

In a drawing of the graph $G\square P_n$, let us consider the following types of possible crossings on the edges of $Q^{(i)}$ for all $i = 1, 2, ..., n-1$:

(1) a crossing of an edge in $G^{(i-1)} \cup M^{(i-1)}$ with an edge in $G^{(i+1)} \cup M^{(i)}$,

(2) a crossing of an edge in $M^{(i-1)} \cup M^{(i)}$ with an edge in $G^{(i)}$,

(3) a crossing among the edges of $G^{(i)}$.

It is readily seen that every crossing of all types (1), (2) and (3) appears in a good drawing of the graph $G\square P_n$ only on the edges of the subgraph $Q^{(i)}$. In a good drawing of $G\square P_n$, we define the force $f(Q^{(i)})$ of $Q^{(i)}$ in the following way: every crossing of type (1), (2) and (3) contributes the value 1 to $f(Q^{(i)})$. The total force of the drawing is the sum of $f(Q^{(i)})$. It is easy to see that the number of crossings in the drawing is not less than the total force of the drawing. The aim of this proof is to show that if every of the subgraphs $G^{(i)}$, $i = 0, 1, 2, ..., n$, has at most one crossing on its edges, then $f(Q^{(i)}) \geq 2$ for all $i = 2, 3, ..., n-2$ and $f(Q^{(i)}) \geq 1$ for $i = 1$ and $i = n-1$.

Consider now the good drawing D of $G\square P_n$ assumed in Lemma 1. Since the graph $Q^{(i)}$ contains $S_3\square P_2$ as a subgraph, the graph $Q^{(i)}$ has at least one crossing. The cycles $C_4^{(i)}$, $C_4^{(i-1)}$ and $C_4^{(i+1)}$ can not cross each other. Because if two cycles cross each other, they cross each other at least two times and this contradicts the assumption that every of the subgraphs $G^{(i)}$ has at most one crossing on its edges. Moreover, none of the 4-cycles separates two other. Otherwise in case when $C_4^{(i-1)}$ separates $C_4^{(i)}$ and $C_4^{(i+1)}$ or $C_4^{(i+1)}$ separates $C_4^{(i)}$ and $C_4^{(i-1)}$, the cycle $C_4^{(i-1)}$ or $C_4^{(i+1)}$ is crossed by all four edges joining the separated 4-cycles. In case when $C_4^{(i)}$ separates $C_4^{(i-1)}$ and $C_4^{(i+1)}$, for all $i = 2, 3, ..., n-2$, the path $d_{i-1}e_{i-1}e_ie_{i+1}d_{i+1}$ crosses the cycle $C_4^{(i)}$. Then there is one another crossing between an edge in $C_4^{(i)}$ and an edge of the path $d_{(i-1)}d_{(i-2)}e_{(i-2)}f_{(i-2)}f_{(i-1)}f_if_{(i+1)}f_{(i+2)}e_{(i+2)}d_{(i+2)}d_{(i+1)}$. This contradicts the assumption that every of the subgraphs $G^{(i)}$ has at most one crossing on its edges.

First we consider that the cycle $C_4^{(i)}$ has an internal crossing. This crossing is of type (3) and so, $f(Q^{(i)}) \geq 1$. The assumption that every of the subgraphs $G^{(i)}$ has at most one crossing on its edges implies that $\mathrm{cr}_D(G^{(i)}, M^{(i)} \cup G^{(i+1)}) = 0$ and $\mathrm{cr}_D(G^{(i)}, M^{(i-1)} \cup G^{(i-1)}) = 0$. As the 4-cycles can not cross each other and none of the 4-cycles separates two other, either the cycle $C_4^{(i+1)}$ crosses the $M^{(i-1)}$ or the subgraph $C_4^{(i+1)}$ must be located in a region where the border contains at most two vertices of the cycle $C_4^{(i)}$. The subgraph induced on the vertices $C_4^{(i-1)}$ and $C_4^{(i)}$ except the edges $\{a_{i-1}d_{i-1}\}$ and $\{a_id_i\}$ is 2-connected graph. So, the cycle $C_4^{(i+1)}$ crosses the subgraph $M^{(i-1)}$ at least two times. It implies, that in both cases the subgraph $C_4^{(i+1)} \cup M^{(i)}$ crosses the subgraph $C_4^{(i-1)} \cup M^{(i-1)}$ at least two times and $f(Q^{(i)}) \geq 3$. Hence, the cycle $C_4^{(i)}$ has no internal crossing. It implies from the assumption that at most one of the subgraph $C_4^{(i+1)} \cup M^{(i)}$ or $C_4^{(i-1)} \cup M^{(i-1)}$ crosses the cycle $C_4^{(i)}$. Without loss of generality let $\mathrm{cr}_D(C_4^{(i-1)} \cup M^{(i-1)}, C_4^{(i)}) = 0$. As the 4-cycles can not cross each other and none of the 4-cycles separates two other, either the cycle $C_4^{(i+1)}$ crosses the $M^{(i-1)}$ or the subgraph $C_4^{(i+1)}$ must be located in a region where the border contains at most two vertices of the cycle $C_4^{(i)}$, again. Then the same analysis as described above implies that $f(Q^{(i)}) \geq 2$.

It remains to prove that $f(Q^{(i)}) \geq 1$ for $i = 1$ and $i = n-1$. First we prove that $f(Q^{(1)}) \geq 1$. Any drawing of $K_{3,3}$ contains a pair of edges that cross each other and do not meet in a vertex. Let us denote by $Q_K^{(1)}$ the subgraph of $Q^{(1)}$ obtained from $Q^{(1)}$ by removing six vertices b_j and f_j for $j = 0, 1, 2$ and two edges $\{d_0, d_1\}$ and $\{d_1, d_2\}$. The graph $Q_K^{(1)}$ can be obtained by an elementary subdivision of six edges of $K_{3,3}$. So, there is a forced crossing between an edge in $G^{(0)} \cup M^{(0)}$ and an edge in $G^{(2)} \cup M^{(1)}$, or between an edge in $G^{(1)}$ and an edge in $M^{(0)} \cup M^{(1)}$, or between an edge in $G^{(1)}$ and an edge in $G^{(0)} \cup G^{(2)}$. Every crossing among the edges of $G^{(0)} \cup M^{(0)} \cup G^{(1)}$ and every crossing between an edge in $G^{(0)} \cup M^{(0)} \cup G^{(1)}$ and an edge in $G^{(2)} \cup M^{(1)}$ are contained only in the subgraph $Q^{(1)}$. So, it contributes the value 1 to $f(Q^{(1)})$. It implies that possible crossing is between an edge in $G^{(1)}$ and an edge in $M^{(1)}$. This crossing is the crossing of type (2) and it contributes the value 1 to $f(Q^{(1)})$. So, we get $f(Q^{(1)}) \geq 1$, again. The same analysis can be repeated for the subgraph $Q^{(n-1)}$ and this completes the proof.

□

The next theorem determines the crossing number of the graph $G\square P_n$ for $n \geq 3$.

Theorem 1 $\mathrm{cr}(G\square P_n) = 2(n-2)$ *for* $n \geq 3$.

Proof. The drawing in Fig. 2 shows that $\mathrm{cr}(G\square P_n) \leq 2(n-2)$, because every copy of $G^{(i)}$, $i = 2, 3, ..., n-2$, is crossed two times, $G^{(1)}$ and $G^{(n-1)}$ are crossed once and there is no other crossings in the drawing. We prove the reverse

inequality by the induction on n. Since the graph $S_3\Box P_3$ is a subgraph of $G\Box P_3$ and we know that $\mathrm{cr}(S_3\Box P_3) = 2$ (see [3]), the crossing number of $G\Box P_3$ is at least two. The reverse inequality $\mathrm{cr}(G\Box P_3) \leq 2$ follows from the drawing in Fig. 2. Assume that $\mathrm{cr}(G\Box P_n) \geq 2(n-2)$ for all integer $n \leq k$, $k \geq 4$, and suppose that there is a good drawing of $G\Box P_{k+1}$ with fewer than $2(k-1)$ crossings. By Lemma 1, some of the subgraphs $G^{(i)}$, $i = 0, 1, ..., k$, must be crossed at least twice. If $G^{(0)} \cup M^{(0)}$ has at least two crossings on its edges, then deleting of all vertices of $G^{(0)}$ results in a drawing of the graph $G\Box P_k$ with fewer than $2(k-2)$ crossings. This contradicts the induction hypothesis. The same contradiction is obtained, if at least two crossings appear on the edges of $M^{(k)} \cup G^{(k+1)}$. If some $G^{(i)}$, $i = 1, 2, ..., k$, is crossed at least twice, by the removal of all edges of this $G^{(i)}$, a subdivision of $G\Box P_k$ with fewer than $2(k-2)$ crossings is obtained. This contradiction with the induction hypothesis completes the proof. □

III. Conclusion

Computing the exact value of crossing number of a given graph is in general an elusive problem. In this paper, we determined crossing number of Cartesian product $G\Box P_n$ for one special graph G of order six. There are known the crossing numbers of products $G\Box P_n$ for many 6-vertex graphs G, but it is still not determined the crossing numbers of all graphs of order six with paths.

Acknowledgment

The author thanks doc. RNDr. Marián Klešč, PhD. for help and precious advices.

References

[1] L. W. Beineke, R. D. Ringeisen, On the crossing numbers of products of cycles and graphs of order four, J. Graph Theory, 4, 1980, 145–155.
[2] D. Bokal, On the crossing number of Cartesian products with paths, J. Combin. Theory (B), 97, 2007, 381–384.
[3] S. Jendroľ, M. Ščerbová, On the crossing numbers of $S_m \times P_n$ and $S_m \times C_n$, Časopis pro pěstování matematiky, 107, 1982, 225–230.
[4] M. Klešč, The crossing numbers of Cartesian products of stars and paths or cycles, Mathematica Slovaca, 41, 1991, 113–120.
[5] M. Klešč, The crossing numbers of products of paths and stars with 4-vertex graphs, J. Graph Theory, 18, 1994, 605–614.
[6] M. Klešč, The crossing number of $K_{2,3} \times P_n$ and $K_{2,3} \times S_n$, Tatra Mt. Math. Publ., 9, 1996, 51–56.
[7] M. Klešč, The crossing numbers of products of 4-vertex graphs with paths and cycles, Discuss. Math Graph Theory, 19, 1999, 59–669.
[8] M. Klešč, The crossing numbers of Cartesian products of paths with 5-vertex graphs, Discrete Math., 233, 2001, 353–359.
[9] M. Klešč, J. Petrillová, The crossing numbers of products of path with graphs of order six, Discuss. Math Graph Theory (to appear).
[10] D. Kravecová, The crossing number of $P_5^2 \times P_n$, Creative Mathematics and Informatics, 21, 1, 2012, 65–72.
[11] Y. H. Peng, Y. C. Yiew, The crossing number of $P(3,1) \times P_n$, Discr. math., 306, 2006, 1941–1946.
[12] J. Wang, Y. Huang, The crossing number of $K_{2,4} \times P_n$, Acta Math. Sci., Ser A, Chin Ed., 28, 2008, 251–255.
[13] L. Zhao, W. He, Y. Liu, X. Ren, The crossing number of two Cartesian products, International J. Math. Combin., 1, 2007, 120–127.
[14] W. Zheng, X. Lin, Y. Yang, Ch. Cui, On the crossing number of $K_m\Box P_n$, Graphs Combin., 23, 2007, 327–336.

Towards Cloud Robotics Age

[1]*Daniel LORENČÍK (1st year)*
Supervisor: [2]Peter SINČÁK

[1,2]Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

[1]daniel.lorencik@tuke.sk, [2]peter.sincak@tuke.sk

***Abstract*—This paper sums the preliminary work done towards the midterm thesis Cloud robotics. The result of the midterm thesis should be architecture of cloud based system providing the tools of artificial intelligence as a service on demand with the use of the internet connection. The proposed solution is to use public cloud computing services as they provide cost effective environment to run applications. In this paper, some of solutions provided will be discussed, and the cloud system will be proposed.**

***Keywords*—cloud computing, cloud robotics, Windows Azure, Amazon Web Services, Infrastructure as a service, software as a service, platform as a service**

I. Introduction

The cloud computing is considered a new paradigm in programming. This may not be entirely true as the cloud computing has many similarities to the grid, from which it has evolved, and also draws concepts from utility computing, service computing and distributed computing [1]. The illustration of relationship of cloud computing and other means of distributed or large scale computing is on the Fig. 1. In [1] the throughout comparison of clouds can be found, main difference is that the cloud is a service oriented versus the application oriented grids. While grids use virtualization, the organization providing the computers to the grid can manipulate the underlying architecture and have control of their assets in the grid. Also, the grids define and provide the set of standard protocols, middleware, toolkits, and services.

The clouds are provided on the per-usage pricing premises, usually from third party companies with datacenters scattered around the world to provide global connectivity with small latency. Clouds provides three types of services:

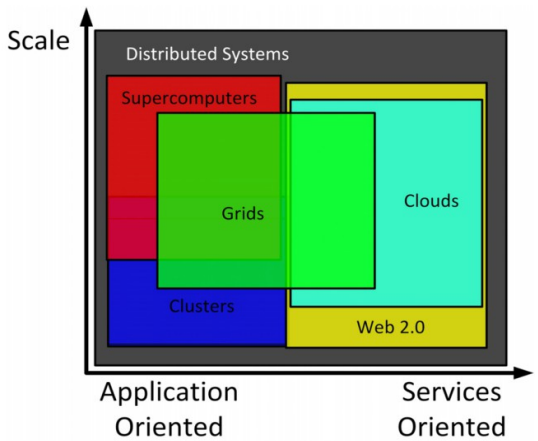


Fig. 1: Comparison of several types of computing depending on the orientation and scale [1]

Infrastructure as a Service (IaaS) provide the user with the virtual machine created depending on the needs, and the cost of the virtual machine depends on the parameters. Infrastructure can be scaled up or down depending on the current need.

Platform as a Service (PaaS) provide the user with high-level integrated environment to build, test and deploy applications. Depending on the provider, there may be some restrictions to the programming software in return for built-in scalability.

Software as a Service (SaaS) provide the user with the possibility to deploy software to end-users through the Internet.

In [2], there is another type of service defined: *Robot as a Service* (RaaS). It is a virtualized agent in the cloud, that can be seamlessly connected to the real robot.

The goal of our research should be cloud based system which will be able to communicate with various types of robots and other devices, and also be able to provide shared storage for the knowledge gained by robots. One part will be to move already developed a system called MASS [3] to the public cloud and provide the plugins from it as a SaaS to end-users, who can be either humans, robots, devices or another service. To this end, overview of cloud computing, existing cloud robotics solutions and commercial public cloud service offers is done.

This paper is structured as follows: In section 2, cloud computing is defined. Section 3 holds the examples of cloud robotics – RoboEarth and DAvinCi. In section 4, two possible options to choose from are described, namely Windows Azure and Amazon Web Services. Section 5 contains the draft of the system we want to implement, and finally the section 6 concludes the paper.

II. Cloud Computing

Formal definition of cloud computing by the National Institute of Standards and Technology (NIST) can be found in [4]. The cloud model is composed of five essential characteristics: *on-demand self-service, broad network access, resource pooling, rapid elasticity, measured service;* three service models: *SaaS, PaaS, IaaS;* and four deployment models: *private cloud, community cloud, public cloud, hybrid cloud.* The clouds rise to the prominence of everyday computing because it also brings new business model, which

is one of the five essential characteristic – *measured service,* or pay-for-use. In section 4, two providers of public cloud services will be discussed along with the detail of how the service is measured.

In real life usage, the *measured service* means that the customer is paying for the resources consumed, similar to the paying system for electricity. Each service model has its own method for measure.

Another important factor with cloud is the *rapid elasticity*, which solve the problem if the computational power or the storage need to be increased, or decreased. Depending on the service model, resources can be changed by the system administrator or developer (IaaS), or they can be changed by the system automatically (PaaS, SaaS).

From the security and legal point of view, the resource pooling is important as generally it is not possible to have control or exact location of the provided resources, therefore to know, where are data stored. It may be possible to know or define the location of data on higher abstraction layer as is a country, or datacenter.

Deployment modes are rather self-explaining. As the results of our research should be also services available for public use, we will focus on the public cloud and public cloud solution providers. In all deployment models, the cloud can communicate with institution datacenter or servers, and act as a part of internal network.

Service models has been already briefly described in section 1.

The access to the cloud in every service is done usually with the use of the web interface, or specialized software using known protocols as *HTTP, HTTPS* and else.

In [5], cloud computing is divided to two categories based on what the cloud provide – *instances on demand* and *computing on demand.* The IaaS deployment mode is from the first category, the PaaS and SaaS from the second.

The MapReduce algorithm [6] is representative of the *computing on demand* option, which is used on large data sets to allow for the parallel processing of the data.

Coupled with cloud computing is the cloud storage. This can also be the service on demand. This couple can be denoted as a *cloud*.

But the usage of cloud services requires a change in the development of applications. From end-user view, the managing the updates and patches of a cloud based application provided as a service is no longer her concern – the new version or patch is in use as soon as it is finished and uploaded. This also can require the change in business model of such applications to be also paid-per-use.

III. Cloud Robotics

Cloud robotics is an application of clouds with a focus on robots. The main idea is to share the resources between robots as in the cloud. The main goal of RoboEarth [7] group is to create a *Rapyuta* [8] – system that will allow to connect different robots to the central knowledge base. The benefits of this system would be instantaneous sharing of the acquired knowledge. Lets say, if one robot learns a new action (how to do something), or is shown some new object for interaction (and how to interact), it can store this new knowledge in the database, where it can be accessed by other robot connected to it. That way, the robots become smarter. The cloud is used as a repository in this case, the decision process and execution of tasks are done on the robot itself.

Another way how to use the cloud in robotics is described in DAvinCi framework [9], where Hadoop, open source implementation of MapReduce [10] is used. The system consists of several robots able to navigate in the environment and at the same time capturing the map of the environment. The data from the robots are stored centrally and used to build the map from the partial maps provided by the robots. Also, the system coordinates the robots so that the parts of the environment already mapped are not traversed again. In the following subsections is the more detailed description of the both systems [11].

A. RoboEarth engine / Rapyuta

The main goal of RoboEarth / Rapyuta [7], [8] is to create a network for robots which will allow robots to exchange knowledge, will have plug-in architecture, and will use ontology-based language. There are three databases on the cloud which hold information about actions, about objects and about environments. Data in these databases are stored in OWL ontology language. Connection to the robot is done by robotic operating system (ROS) [12], but the system itself is not limited to the use of ROS for communication. The information about robot capabilities (construction, sensor types and others) are published for the system, so only the appropriate actions (the actions the robot can do) are provided. Main component of RoboEarth architecture is Recognition/Labeling component (RLC). It connects robot hardware with the abstractions of actions, objects and environments. Its main function is to translate abstract definitions from RoboEarth databases to format understood by particular robot and vice-versa so the robot can contribute to the databases with new knowledge. That means it is able to work on low-level actions (atomic primitives – signals from sensors, motors, and others) and also on high-level actions (spatial and temporal relations between actions to create and execute action abstractions).

The execution of task begins with downloading the appropriate action recipe for the task. This is then transformed into finite state machine by the execution engine, and the plan is executed by traversing the nodes of this machine, so the nodes are mapped to trigger the actions for the robot hardware.

Learning the new action is done through teleoperation. Human operator controls the robot, and since the RoboEarth component is on the ROS blackboard, it receives all the signals from the sensors and the motors, and wraps them into the action plan. Then it asks the operator for the action label. The operation of learning is also running when robot is executing the action recipe – to further improve it.

B. DAvinCi framework

DAvinCi framework [9] is a software framework that provides the scalability and parallelism advantages of cloud computing for service robots in large environments. For communication with the robots, the ROS is used. For parallel processing, the Hadoop MapReduce framework [10] was

used. The framework capabilities were tested on the implementation of FastSLAM algorithm to build a map of large arena using different robots. DAvinCi is a Platform as a Service: robots communicate with the DAvinCi server, which can run ROS nodes for the robot without the capability to run them (Roomba, Rovio). Hierarchically, above the server is located the Hadoop Distributed File System (HDFS) cluster used for execution of robotic algorithms. Also, DAvinCi server acts as a central communication point, as well as the master node. On this framework, as a proof of concept the FastSLAM algorithm was implemented and tested with promising results.

IV. Public Cloud Solutions

There are several public cloud solutions providers, which differ in the offers. But the same is the pay-per-use pricing model. Our research will require the use of several programming languages and will benefit from fewer restrictions. For the final decision between providers, we had chosen Windows Azure from Microsoft and Amazon Web Services. Other providers (AppFog, Ciright Systems, Cloud Foundry, Engine Yard, Heroku, Mendix, OpenShift, and others) were ruled out because they are either in the development stage, or do not provide required support for .Net written applications, or they do not own global network of interconnected datacenters. Google API Engine provides PaaS which could be suitable, but it supports only Python, Java and Go. Although the platform is build as language independent, at the time of writing this paper only these (and by extension others JVM languages) are supported. We do not abandon the idea of using the Google App Engine in the future, or other possibilities that can emerge, but currently we are focusing on the aforementioned solutions. The overview of the chosen solutions will be given in following subsections.

A. Windows Azure

Windows Azure is the public cloud solution provided by Microsoft [13]. The main advantage is the support of a wide array of programming languages (Java, C#, Node, PHP, Python). Also, there are an extensive documentation (in MSDN Library) and courses available. The available services are grouped into categories shown on the Fig. 2.

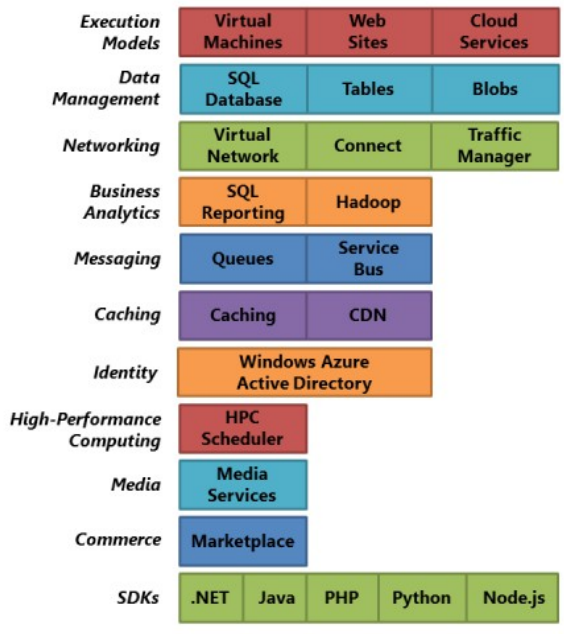


Fig. 2: Windows Azure services categories [19]

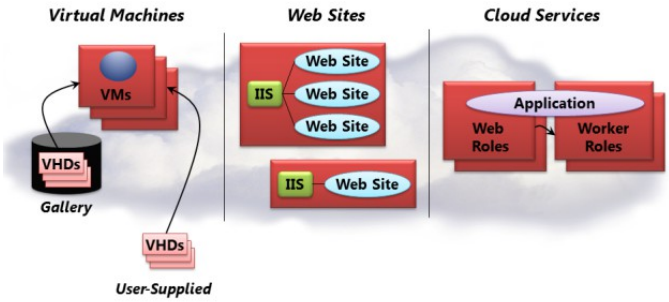


Fig. 3: Windows Azure execution models [19]

On Azure, three execution models [14] are present (illustrated on Fig. 3):

Virtual Machines provides the user with virtual machine (VM), the size of VM can be *extra small* (shared core and 768 MB of memory), *small* (1 core and 1.75 GB of memory), *medium* (2 cores and 3.5 GB of memory), *large* (4 cores and 7 GB of memory), *extra large* (8 cores and 14 GB of memory). On these VMs, the operating system can be installed from the gallery of virtual hard disks (VHD) provided by Microsoft and its partners, or the customer can provide her own image of VM. *Virtual Machines* correspond to the IaaS.

Web Sites run on the virtual machines with the Windows Server and Internet Information Services (IIS) and is similar to the classic web hosting service. *Web Sites* allow to host the simple static websites as well as the full-fledged web applications created by using ASP.NET, PHP, Node.js, which can use other parts of Azure (service bus, SQL database Storage).

Cloud Services provide PaaS. The technology used is supporting scalable, reliable applications. Applications are run on the virtual machines of two types. *Web roles* run on the VM with Windows Server with IIS, and the *Worker roles* run on the VM with Windows Server without IIS. The management of VM is done by the Azure. This is the difference between PaaS and IaaS. In the configuration file of cloud application, the developer sets the number of needed instances of VM, which are created on the upload of the application to the Azure.

The *Storage* options are relational SQL-like databases (SQL Azure), Tables and Blobs. SQL Azure (or only SQL database) is cloud version of SQL. The underlying system creates the replicas automatically and reroutes the request to the replica if an error with the database occurs. It is also possible to use the full version of SQL server on VM and connect it to the cloud service, but the system will not be able to manage it automatically. Tables are used to store non-related unstructured data in key-value pairs. This service can autoscale to house 100 terabytes of data [15]. Blobs (Binary Large Object) are used to store large amounts of unstructured text or binary data (images, video, audio). It is possible to add structured data to the blob in the form of metadata or it can be mounted by the application as a single volume NTFS VM.

B. Amazon Web Service

Amazon Web Service (AWS) provides similar products to Windows Azure, and the largest network of datacenters. There is plenty of documentation [16].

Amazon Elastic Compute Cloud (EC2) is similar to the *Virtual Machines* in Azure as it provides the user with the virtual machines according to the needs. Compared to the Azure, there are much more options available (complete list on [17]).

Storage options similar to blobs in Azure are *Simple Storage Service (S3), Glacier, Elastic Block Store (EBS). S3* allows to store data in objects accessible via unique developer-assigned keys. Glacier is a persistent storage optimized for storing backups, archives or data that are accessed infrequently. *EBS* are storage volumes that can be attached to the EC2 instance and store database, file system, or raw data. The databases are managed through the *Relational Database Service (RDS),* and non-SQL databases are *DynamoDB, SimpleDB.*

PaaS is provided under the name *AWS Elastic Beanstalk*, currently in beta stage [18]. *Elastic Beanstalk* utilizes *EC2, S3* and *Amazon Simple Notification Service (SNS)*. It allows to run Apache HTTP server for Node.js, PHP and Python, Passenger for Ruby, IIS7.5 for .Net and Apache Tomcat for Java. By default, the application is running on single *EC2* micro instance, and in the case of peak in workload or traffic, the elastic load balancer will create and run additional instances, or shut them down accordingly.

V. Proposed System

As was already mentioned, the system we want to create is based on the MASS [3], and it should provide the current functionality as a service available through the Internet. To do that, we need to utilize existing PaaS system. The reason why to use public instead of private cloud is the scalability and cost-effectiveness. As the administration and underlying structure are provided, we can instead focus on the applications.

The another con of using the cloud is that if the application we provide as a service is updated, it will be updated immediately for all users. Users could be humans, which need some of the tools in their own programs, or it could be used in teaching as a demonstration tool, or it can be used by robots in their operational programs. What more, the result of one application can be the parameters for others (for example multiple image filters or processing), which can increase the effectiveness of programming for robots.

It can also grant immense computational power to otherwise small and resource-restricted robots (smartphone robots, drones, etc), which can then use more complex algorithms to operate.

VI. Conclusion

We described two major cloud services providers in this paper. Based on the research of the offers and our needs, we have chosen Windows Azure as our PaaS provider because:

1. Extensive documentation
2. C# and .Net support, along with open-source options
3. Level of integration with Microsoft other products
4. Price-perfomance ratio

Not in the least was our decision influenced by the Microsoft support of our faculty and availability of courses provided by Microsoft for the Azure developers.

We may use other provider in the future in case our needs or the offer will change.

Research goal of thesis is the detailed proposal of system (refine the preliminary proposal from section 5), and study of cloud robotics, cloud based image recognition, and how could the use of the cloud based system influence the core method of artificial intelligence.

Technological goal of thesis will be the creation of proposed system and its implementation on the robots available in the lab.

Acknowledgment

This work is the result of the project implementation: Development of the Center of Information and Communication Technologies for Knowledge Systems (ITMS project code: 26220120030) supported by the Research & Development Operational Program funded by the ERDF and by the National Research and Development Project Grant 1/0667/12 "Incremental Learning Methods for Intelligent Systems" 2012-2015.

References

[1] I. T. Foster, Y. Zhao, I. Raicu, and L. Shiyong, "Cloud Computing and Grid Computing 360-Degree Compared," in *2008 Grid Computing Environments Workshop*, 2008, pp. 1–10.

[2] Y. Chen, Z. Du, and M. García-Acosta, "Robot as a Service in Cloud Computing," *2010 Fifth IEEE International Symposium on Service Oriented System Engineering*, pp. 151–158, Jun. 2010.

[3] T. Reiff and P. Sinčák, "Multi-Agent Sophisticated System for Intelligent Technologies," in *International Conferencer on Computational Cybernetics*, 2008, pp. 37–40.

[4] P. Mell and T. Grance, "The NIST Definition of Cloud Computing Recommendations of the National Institute of Standards and Technology," *Nist Special Publication*, vol. 145, p. 7, 2011.

[5] R. L. Grossman, "The Case for Cloud Computing," *IT Professional*, vol. 11, no. 2, pp. 23–27, Mar. 2009.

[6] J. Dean and S. Ghemawat, "MapReduce : Simplified Data Processing on Large Clusters," *Communications of the ACM*, vol. 51, no. 1, pp. 107–113, 2008.

[7] "RoboEarth Project." [Online]. Available: http://www.roboearth.org/. [Accessed: 12-Oct-2012].

[8] D. Hunziker, M. Gajamohan, M. Waibel, and R. D'Andrea, "Rapyuta: The RoboEarth Cloud Engine," in *Proc. IEEE Int. Conf. on Robotics and Automation (ICRA)*, 2013.

[9] R. Arumugam, V. R. Enti, K. Baskaran, and a S. Kumar, "DAvinCi: A cloud computing framework for service robots," in *2010 IEEE International Conference on Robotics and Automation*, 2010, pp. 3084–3089.

[10] "Apache Hadoop." [Online]. Available: http://hadoop.apache.org/.

[11] D. Lorencik and P. Sincak, "Cloud Robotics: Current trends and possible use as a service," in *SAMI*, 2013.

[12] "The Robotic Operating System." [Online]. Available: http://www.ros.org/wiki/.

[13] "Windows Azure." [Online]. Available: http://www.windowsazure.com/en-us/.

[14] D. Chappell, "Windows Azure Execution Models." [Online]. Available: http://www.windowsazure.com/en-us/develop/net/fundamentals/compute/. [Accessed: 11-Mar-2013].

[15] J. Giardino, J. Haridas, and B. Calder, "How to get most out of Windows Azure Tables." [Online]. Available: http://blogs.msdn.com/b/windowsazurestorage/archive/2010/11/06/how-to-get-most-out-of-windows-azure-tables.aspx.

[16] "AWS Cloud Computing Whitepapers." [Online]. Available: http://aws.amazon.com/whitepapers/.

[17] "Amazon Elastic Cloud Computing." [Online]. Available: http://aws.amazon.com/ec2/.

[18] "AWS Elastic Beanstalk." [Online]. Available: http://aws.amazon.com/elasticbeanstalk/.

[19] "Introducing Windows Azure." [Online]. Available: http://www.windowsazure.com/en-us/develop/net/fundamentals/intro-to-windows-azure/. [Accessed: 11-Mar-2013].

[20] "What is Amazon Web Services?" [Online]. Available: http://docs.aws.amazon.com/gettingstarted/latest/awsgsg-intro/intro.html.

Section:
Informatics & Telecommunications
Poster form

A MATLAB educational software tool for Knowledge Discovery course

[1]*Alexandra LUKÁČOVÁ (2nd year), Lenka MIŠENKOVÁ*
Supervisor: [2]Ján PARALIČ

Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

[1]alexandra.lukacova@tuke.sk, [2]jan.paralic@tuke.sk

***Abstract*—The direct consequence of recent fast paced development in informational and communications technology is that the computers have become very important factor in our society. To a certain extend the education also reflects this recent developments. Because people have different ways of learning, some can assimilate in a better way the knowledge received visually, another auditory. The tool presented here was developed in MATLAB and is consisted of two modules explaining the simple and multiple linear regression topics. The tool has been specifically designed for students attending Knowledge Discovery course to help them better understand the curriculum at particular lesson by visualizing the learning process as well as the influence of algorithm parameters.**

***Keywords*— education, software tool, linear regression, knowledge discovery**

I. INTRODUCTION

The advent of different applications and tools using information and communications technology has revolutionized teaching and learning in higher education. A conclusion could be made that the usage of educational software can greatly improve teaching and learning processes in any given subject and in any institution concerned with higher education. Students of mentioned institutions have the possibility to maximize their learning potential if the information technologies are employed in the support of teaching practices. Additionally the trained individuals gain independence and have the possibility to be more self regulated while learning. This means that information and communication technologies should adapt to easily to different characteristics of students if possible [1], [2].

This paper is organized as follows: Section 2 explains the motivation of developing such tools and their need by students learning, especially in Knowledge Discovery course. Section 3 presents the tool that has been developed to motivate students in learning and to help them better understand linear regression problems by visualization of the learning process. Finally, conclusion and student feedback is presented together with ideas for future work.

II. RELATED WORK

There exist many software tools, which were exclusively developed for the different lesson needs as an aid for enhancing students learning. Anyway good tool, if used appropriately, can enhance student performance. Some of typical model examples are listed below.

To assist students in improving their problem solving skills and to encourage concept integration in the first-year courses a computational MATLAB Tool for first-year chemical engineering students has been developed. Feedback from students who used the MATLAB Tool during their term through the study group indicated that students recognize the interconnectedness of the course material in the first-year courses [3].

Using packet tracer as a tool in teaching complex and abstract units like Data communication and Networking, Distributed networks significantly helped the students to improve their learning and creativity skills. Skills in problem solving, designing and troubleshooting have greatly improved when students used the CISCO packet tracer. Furthermore, teaching and learning technology became more comprehensive, enjoyable, stimulant when it was used in simulation based class room teaching [4].

Third example represents usage of the program visualization software Jeliot 3 in the Java programming course. This research examined the beginning of the process of comprehension of computer programs in group of novice programmers. In the study, 45 novice programmers were tested in how they comprehend segments of short programs. The results suggested that students who learned with Jeliot 3 made detailed, concrete mental representations of the program text, supporting it with better test examples than students from the control group. Also, they scored better on the test [5].

III. MOTIVATION

Focus of this study is on data-oriented knowledge discovery techniques (KDD – Knowledge Discovery in Databases, or currently often named Business Intelligence or Business Analytics), which represent an iterative semi-automatic process of discovery and verification of the patterns in a large amount of data, using pattern recognition technologies and statistical and mathematical techniques [6]. Data mining, method associated with KDD, was originally used mainly as a solution for marketing departments to achieve increased sales, but now is becoming more widespread and is successfully used also in other fields such as finance, economics and business problems.

Nowadays, a lot of general data mining tools are available

to users. Some examples of commercial mining tools are SAS Enterprise Miner or SPSS Process Modeler. Some examples of public domain mining tools are Weka or RapidMiner. However, all these tools are not specifically designed for educational purposes.

Many of the algorithms covered in Knowledge discovery course involve a lot of visual understanding and are difficult to understand from classroom lectures and textbook reading. Therefore the development of educational software tools is needed.

There is a frequent problem caused by limited time of course, that curriculum at the lesson is not explained sufficiently and student does not understand it adequately. In that case he or she has to make an effort to learn it by him or herself. In this case he or she has to go through different materials, whether scripts, collections of different tasks or solving problems. The disadvantage of textbooks and collections of tasks is that they often contain examples that have already been presented several times. If you take as an example the calculation algorithms of data mining, we observed that a suitable complement to supplementing sources of knowledge for student could also be an application that will simulate the steps of the calculation.

To ascertain the usability of such tools, a questionnaire was developed. The questionnaire was applied on group of 40 people, mostly university students in their final year of master study, who already have considerable experience in the use of instructional support applications and educational tools. Respondents could choose from several options, or they could enter their own answer.

The first question of the questionnaire determines the extent to which people use highly sophisticated educational applications. 68% of respondents have already used an

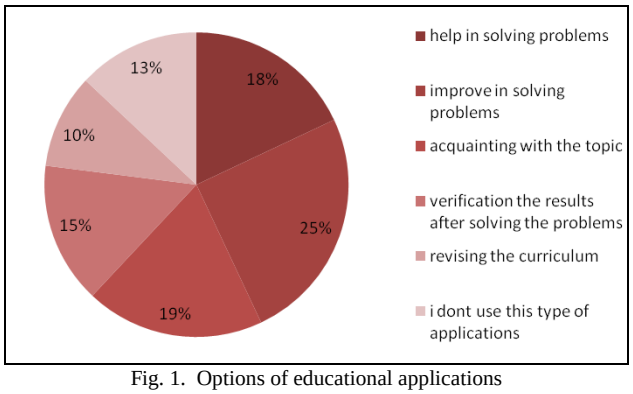


Fig. 1. Options of educational applications

application educationally and only 32% of them did not have. We can see that the most of the respondents have had experience with such applications and with the usage of them. But in spite of the fact that 32% of respondents were not used educationally applications, 95% of respondents expressed will to use them in the future.

The most students (25%) stated they wanted to verify their results after solving the problem by themselves. Additional detail of choosen options for educational applications are presented in the Fig. 1.

Participants of the questionnaire would recommend a tool that has detailed explanation of the problem 16%, students mentioned also simple explanation of the problem 20%, another expected features of educational applications were visualization algorithm steps 20%, helping description of the steps the algorithm 13%, one simple sample input 9%, more sample inputs 10% and 9% students also suggested the ability to add their own input algorithm.

As the most frequent flaws of educational applications students indicated lack of transparency 16%, availability of the requested calculations 20%, inadequate explanation of the problem 17%, complicated explanation of the problem 6%, and too detailed explanation of the problem 17%, lack of quality GUI 24%.

IV. MATLAB-BASED EDUCATIONAL TOOL

One of the topics explaining during the Knowledge Discovery course is simple and multiple linear regression. Regression analysis is a statistical technique for investigating and modeling the relationship between variables [7].

This paper presents the tool exclusively developed in MATLAB for illustrating the applications of the theory taught at the lesson – simple and multiple linear regression. The aim of linear regression, shown in Eq. (1),

$$h_\theta(x) = \theta_0 + \theta_1 x, \tag{1}$$

is to determine parameter values θ_0 and θ_1 in order to give the best possible approximation. To find these values, we use gradient descent algorithm, where the goal is to minimize costs computed by the Eq. (2),

$$J(\theta_0, \theta_1) = \frac{1}{2m}\sum_{i=1}^{m}((\theta_0 + \theta_1 x^{(i)}) - y^{(i)})^2. \tag{2}$$

Based on data collected in previous questionnaire, an application was developed in a way to reflect requested qualities. The big plus of this application is that samples to the base problems are included in the tool to give students the opportunity to apply what they have just learned and problem-solve independently.

MATLAB is a high-level language and interactive environment for numerical computation, visualization, and programming and many engineers and scientists in industry and academia use this language of technical computing [8].

The tool consists of two parts, where first is focused on

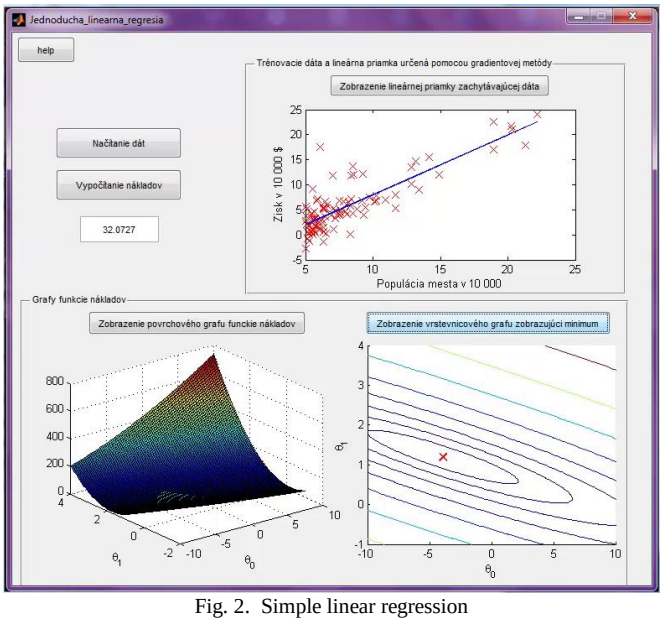


Fig. 2. Simple linear regression

explaining simple linear regression. We helped with MATLAB tutorial topics when creating the application [9]. Whereas the application is designed primarily for Slovak students, print screens showed in this paper are given in Slovak language. As we can see in the Fig. 2, it is possible for student to retrieve his own data by "Read data" button or to choose a dataset from offered samples. On the left side we can see the scatter plot, where the *x* axis represents independent variable and the *y* axis dependent variable. The main goal for

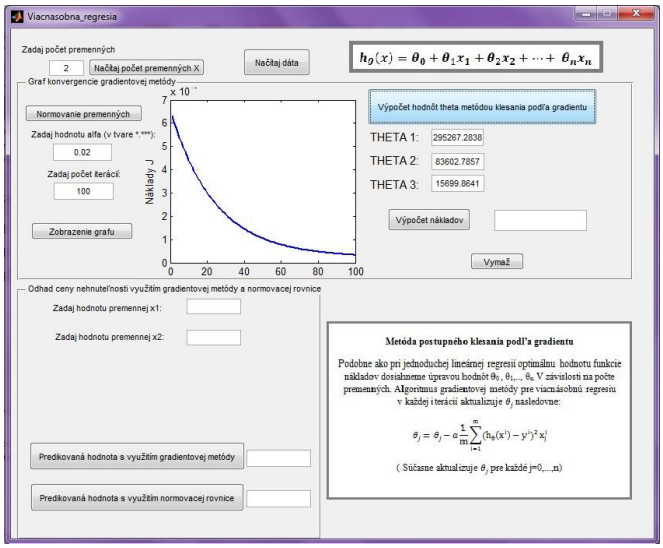


Fig. 3. Multiple linear regression

the student is to find the line, where the costs are minimal. Because of better understanding the tool visualizes the cost function and also wanted parameters - theta 0 and theta 1. For the user there is also the option to try how the line looks like when the parameters will change. For the student there is also always available the help button, if he does not understand the theory. Since datasets included in the application are intended for educationally purposes, they are small. Some of them are inspired by examples of machine learning course, some are created directly by us [7].

Second part of the tool, shown in the Fig. 3, is describing the problem of multiple linear regression. At the beginning it is necessary to keep the sequence of steps. First, enter the number of variables and just after that it is possible to read data from file. As in the first part, there is always help button available and also a window shown the elementary theory.

V. Conclusion and Future Work

In this paper we have presented the educational tool developed for Knowledge Discovery course, which was designed to assist students in improving their ability to use and understand linear regression problems and to encourage them to try it by themselves at home.

We evaluated it on 26 students with questionnaire survey, as it has become the most dominant method of gathering student feedback [10], which followed after the lesson, where was the tool used. Every one of the students considered this type of teaching much more interesting than standard procedures, where the students are just listening to the teacher. At the question, if the application has helped them to better understand the topic, 23 students answered positively and just 3 students were not satisfied. Unsatisfied students declared insufficient help guide as the main reason. As far as understandability is concerned, 16 students mentioned, they had problems with composition of multiple linear regression. This raises our plan to change the help to more intuitive way. From the comparison of the students' performance, who used this tool at the lesson to the students, who do not use it revealed, that students, who had the opportunity to try out the application perceived this teaching options better than students, who never met with such a form teaching yet. The research also shows that students would be interested in this form teaching and they believe that this kind of the application would help them better understand the problem.

Acknowledgment

The work presented in this paper was partially supported by the Cultural and Educational Grant Agency of the Ministry of Education, Science, Research and Sport of the Slovak Republic under grant No. 065TUKE-4/2011 (50%). This work is also the result of the project implementation Development of the Center of Information and Communication Technologies for Knowledge Systems (ITMS 26220120030) supported by the Research & Development Operational Program funded by the ERDF (50%).

VI. References

[1] A. L. Franzoni, S. Assar, B. Defude, and J. Rojas, "Student Learning Styles Adaptation Method Based on Teaching Strategies and Electronic Media," in *Eighth IEEE International Conference on Advanced Learning Technologies, 2008. ICALT '08*, 2008, pp. 778-782.

[2] F. Haq, "The effects of information and communication technology (ICT) on students learning in teacher education programs," in *EDULEARN12 Proceedings*, 2012.

[3] N. L. Cress, M. A. Robinson, L. Corner, R. L. Legge, and L. A. Ricardez-Sandoval, "Problem-solving and concept integration using a computational tool in first-year undergraduate chemical engineering," *Education for Chemical Engineers*, vol. 7, no. 3, pp. 133-138, August 2012.

[4] V. Shanmugam, L. Gopal, Z. Oo, and P. M. Viswanathan, "Enhance student's learning with an aid of simulation software to understand Computer Networking," , 2011.

[5] S. Maravic Čisar, R. Pinter, D. Radoslav, and P. Čisar, "Software visualization: The educational tool to enhance student learning," in *Proceedings of the 33rd International Convention MIPRO*, 2010, pp. 990-994.

[6] J. Paralič, *Objavovanie znalostí v databázach*, 1st ed. Košice: Elfa, 2003.

[7] A. Ng. Stanford Machine learning. [Online]. www.coursera.org/course/ml

[8] Matlab - the language of technical computing. [Online]. http://www.mathworks.com/products/matlab/

[9] Primer. Matlab-R2013a. [Online]. http://www.mathworks.com/help/pdf_doc/matlab/getstart.pdf

[10] A. Coughlan, "Evaluating the Learning Experience: the Case for a Student Feedback System," , February 2004.

A Proposal of Customer Relationship Management System using Distributed Architecture

[1]*Michal ENNERT (2nd year),* [2]*Ján HURTUK*
Supervisor: [3]*Liberios Vokorokos*

[1,2,3]Dept. of Electronics of Computers and Informatics, FEI TU of Košice, Slovak Republic

[1]michal.ennert@tuke.sk, [2]jan.hurtuk@student.tuke.sk, [3]liberios.vokorokos@tuke.sk

***Abstract*—This paper deals with CRM systems with their using in business planning and with their internal structure and differences based on it. Based on previous research, authors decide to create their own CRM system with unique structure using modern approaches connected with variety of Web technologies, Linux based technologies a database systems.**

***Keywords*—CRM, business planning, ETL, database systems**

I. Introduction

CRM (Customer Relationship Management) is common name of systems for customer relationship management. The success of any company is determined by satisfaction of their customers, their satisfaction with products of the company, their satisfaction with access to and willingness to return to the company. That is why every company has to pay adequate attention to customers and time. However, it is very difficult to gather and measure this information and to gain some useful information from it. Implementation of CRM demands systematic approach and represents a matter of organizational design. CRM requires a complex, cross-functional integration of people, processes, operations, and marketing capabilities that is enabled through technology and applications. Many organizations have spent the last twenty years in an attempt to approach the customers by using the buyer-segment methods, "focus group" and by measuring the customer's satisfaction. However, simple placing the buyer in the limelight of the organization does not make the organization customer-oriented. It is required that the organization undergoes reorganization and that it is literally organized around its customers.

Considering the aforementioned, that CRM system represents the main issue for customer-centric companies; below we present the characteristics of structure and processes in an organization with CRM system. Introducing CRM system in considerable number of cases has a positive effect on business results. There is a series of accepted definitions of CRM, and one of the most meaningful was provided by Swift who says that CRM includes efforts of the entire company which are directed towards a better understanding of customer's behaviors and acquiring opportunities to influence such behavior through various forms of meaningful communication, with the purpose of constant improvement of possibility to attract new customers and keeping the old ones, and rising the level of their loyalty and usefulness [1], [2].

II. Structural Analysis of CRM Based System

A. Analyzing of CRM System Architecture

Depending on the concept of understanding CRM, if we observe the problem from tactical perspective, which presents a narrower point of view, then CRM is primarily about the implementation of a specific technology solution project [3]. CRM systems in practice most often include purchase of software and hardware which will enable the company to save important information about certain customers. By studying the past purchases, demography and psychology of a customer, the company gets to know the customer's preferences. In this way, the company can also send specific offers only to those customers with expected high interest for purchase, which brings savings. By using the data carefully, the company can improve attracting the attention of new customers, cross-selling and up-selling [4], [5].

According to relevant analysts, CRM could be classified into several types [3], or, to put it plainly, CRM consists of three components. These three components of CRM support each other and the success of the entire system demands their proper integration:

- Operational CRM-client database,
- Collaborative CRM-contact center, i.e. customer service, web pages designed for customer interaction, communication with clients via available media (e-mail, SMS, telephone, fax, mail, physical contact etc.),
- Analytical CRM-expert in CRM in terms of an employee with required knowledge, and various available CRM systems and applicative solutions.

There are some specific CRM building blocks [6]:

- A database that collects information about your customers.
- A way to analyze the information in the database.
- A strategy for applying the analysis to better meet your clients' needs and identify potential customers.
- Collecting data to ensure your strategy is effective.

B. Usage of CRM in Business Planning

By compiling this information and analyzing it, you can then build a strategy with this information to:

- Maximize repeat business opportunities by anticipating your existing customers' needs.

- Identify potential and best customers.
- Identify complementary products you can sell to your customers.
- Target marketing campaigns/materials and promotions.

There are three basic types of information and requirements obtainable from customer, which is possible to gather by some specific questions as is described in Tab. I.

TABLE I
TYPE OF INFORMATION FROM CUSTOMER

Type of information	**Answer the questions**
Customer profile information	• Who are they? • Are they business or a person? • Where are they located? • If they are business, what do they do? • Why do they need our product? • How do they communicate? • How long have they been a customer?
Customer buying profile	• How often do they buy? • When do they buy? • Is there a pattern to their buying habits? • How much do they buy at one time? Over time?
Customer buying preferences	• What do they buy? • Do they always buy the same thing? • Why do they buy it?

III. IMPLEMENTATION OF CRM SYSTEM PROPOSED MODEL

A. Options for CRM

There is a continuum of CRM from the most simple (a spreadsheet or database containing information about your customers – referred to as a contact management system) to the most complex (online applications automatically linked with your back-end systems). An integrated CRM system can include the following characteristics and benefits described in Tab. II.

TABLE II
CHARACTERISTIC AND BENEFITS OF PROPOSED CRM

Characteristics	**Benefits**
A central database that is accessible by all employees to view and update customer data.	+ Improved customer service, loyalty and retention.
Analysis of customer data including customer segmentation and segmentation of potential customers.	+ Customized marketing or sales campaigns. + Improved campaign targeting.
Customer self-service where the customers can self-order and help themselves using web-based, password access.	+ Reduced order entry cost and customer service cost
Identifying and tracking potential customers.	+ Wider customer base. + More focused prospect tracking.
Reports generated with up-to-date information, including revenue forecasting and trend analysis.	+ Better and timelier decision making.

Based on used technologies exist two main options: A CRM package installed on your premises. There are many CRM systems that are available for purchase off-the-shelf. These can then be tailored to customer needs. Companies such as Siebel, Oracle, SAP, and Chordiant are well known in this area. The second option is *Hosted CRM*. There are web-based applications for CRM with no software to download. In this case, the CRM system resides online and you rent the service on a monthly basis. Examples include Sage Software, Microsoft Dynamics, Entellium, Clear C2, SAP, NetSuite, Vanilla Soft. Advantages and disadvantages of these two options are described in Tab. III.

TABLE III
ADVANTAGES AND DISADVANTAGES OF USED CRM TECHNOLOGIES

Description	**Pros**	**Cons**
On-premise CRM	+ Can be tailored to your business. + Can be integrated with your other systems. + Most companies offer flexible packages that are suitable for small and medium businesses.	- More expensive in the short run (from several thousand to several million dollars/euro). - Can take a long time (months or even years) to implement fully.
Hosted CRM	+ Less expensive in the short run (monthly fees run from about $65 to $150). + Appropriate for businesses with standard CRM needs, and little or no internal IT support. + Can be implemented quickly (often within a few months).	- Cannot be integrated with other back office systems. - Owner is allowing someone else to control customer information and data.

B. Questions to Consider before Using CRM

CRM is a business strategy, not a technology. For CRM to succeed in the company, it is very important to develop the CRM strategy first, and then choose the best technology to support it. When starting to collect the data, there are four important things to consider:

- Has the current system the capacity to hold and manage demanded amount of data?
- Is every piece of collected data really necessary to store?
- Are among stored data some, which need to be encrypted and stored especially carefully because of privacy policy?
- Will be stored data used for company's internal improvement only?

IV. Internal Structure of Proposed Web-Based CRM System

A. General Model of the Proposed System

General model of web based CRM system is designed based on three main parts: Web-based server, Email Server, ETL's. This model is showed on Fig. 1.

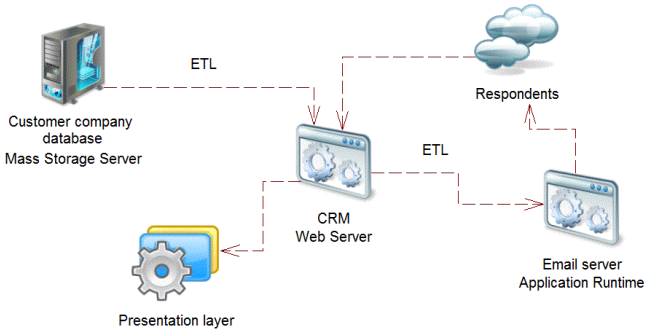


Fig. 1. General Model of Web Based CRM

Web server serves as storage and runtime environment for ETL scripts, for representing collected data and for communicating with email service based server.

Email server serves as runtime environment for sending messages to the customers to collect needed data. It can be represented by service attached to the whole CRM system or by external company specialized to email services.

ETL's (Extract, Transform and Load) are scripts, which join aspects of the represented model together. They are divided into two main parts: First part is the ETL script for gaining customer profile information (date of birth, bought products etc.) and Customer buying profile. Second part contains the ETL script designed for communication with email server, which helps to send surveys with questions about Customer buying preferences and Customer intentions to do business in the future.

B. Model of the Central Web Server

The central web server is divided into four main parts, which are connected in the mesh pattern via ETL's: Presentation layer, Administration layer, Survey layer, Email sending layer. This model is showed on Fig. 2.

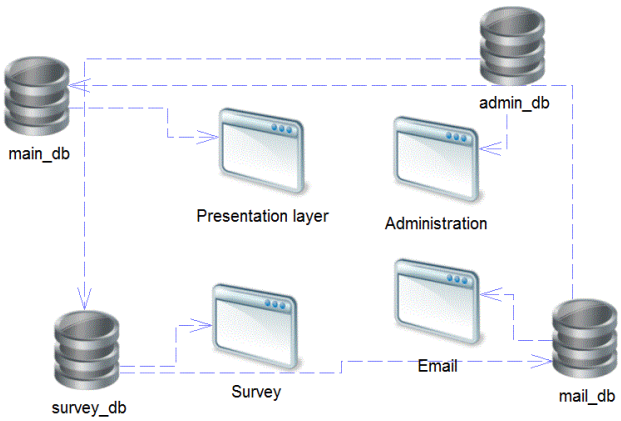


Fig. 2. Model of the Central Web Server

Presentation layer is a web based application, which takes care of representing collected data in appropriate way. It should be designed to represent data in user-friendly manner, which means, that the represented data should be comprehensible even for person with no IT or Mathematical based education.

Administration layer is a web based application, which takes care of configuring questions for survey (questions about Customer buying preferences and Customer intentions to do business). It should be designed user-friendly and be usable even for person without IT based education. Administration layer is important to consider, when the developer wish to hand over the control of the system to the customer company.

Survey layer is web based application used to represent selected questions (questions about Customer buying preferences and Customer intentions to do business).

Email sending layer is responsible for sending an original email to the company customer. Every email should contain an original http hypertext link on survey layer (this link is generated for every registered company customer and has to be unique).

C. Model of the Presentation Layer

Presentation layer can be divided into huge variety of web pages as is described on Fig. 3. It is all based on method of grouping the represented data unions.

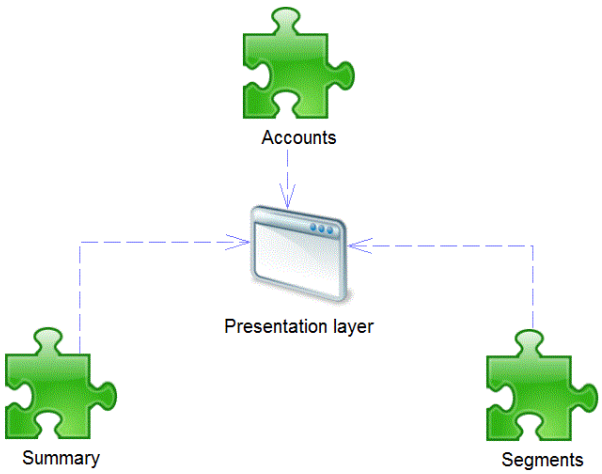


Fig. 3. Presentation Layer Model

Summary page: can represent all data as one union. It can show how many respondents are taking part in the iteration, how many of them would intent to do business in the near future etc.

Segments page: can divide respondents based on age, gender, city or state where they live etc.

Accounts page: can show individual respondents data collected from both sides (both ETL's), from company side and from survey data side.

V. Conclusion

In these days, days of free market and days of great opportunities and variety of products is careful business planning very important for every businessman who wants to be successful. The CRM systems represent an easy way to build successful business strategy based on interaction with customers and based on monitoring customer needs. The great

advantages of proposed model are high scalability for customers, modularity of proposed system and distributed computing performance into parts of architecture.

For future successful implementation of designed CRM into production, it should accomplish some requirements:

- It should be user friendly.
- It should handle a huge amount of data.
- The design pattern should be modular.
- It should be an easy to modify and to implement into various environments.

To achieve these goals, series of tests need to be performed and careful choosing of technologies needs to be done.

ACKNOWLEDGMENT

This work was supported by the Slovak Research and Development Agency under the contract No. APVV-0008-10 and No. APVV-0073-07.

REFERENCES

[1] Swift, R.S.: Accelerating customer relationships: using CRM and relationship technologies. Prentice Hall Professional, 2001.

[2] Dua, S., Sahni, S., Goyal, D. P.: Information Intelligence, Systems, Technilogy and Management. Proceedings of the 5th International Conference, ICISTM 2011, Gurgaon, India, March 10-12, 2011.ISBN 978-3-642-19423-8.

[3] Payne, A.: Handbook of CRM: Achieving Excellence through Customer Management. Butterworth-Heinemann, 2006 p.19, p.24.

[4] Kotler, P.: Marketing insights from A to Z. John Wiley & Sons, Hoboken, New Jersey, 2003 p.35.

[5] Oz, E.: Management Information Systems [With Access Code]. Cengage Learning. 2008. ISBN 1-4239-0178-9.

[6] Customer Relationship Management. How You Can Profit from E-Bussiness. [online], 2010. <http://www.ontariocanada.com/ontcan/1medt/smallbiz/sb_downloads/ebiz_customer_relationship_en.pdf>

[7] Khodakarami, F., Chan, Y.: Evaluating the Success of Customer Relationship Management (CRM) Systems. Proceedings of the 2nd International Conference on Information Management and Evaluation: Icime 2011, Ryerson University, Toronto, Canada, April 27-28 2011. ISBN 97-1-906638-97-9.

[8] Tomášek, M.: Language for a Distributed System of Mobile Agents. Acta Polytechnica Hungarica, Vol. 9, No. 2, ISSN 175-8860, p. 61-79, Budapest, 2011.

[9] Vokorokos, L., Baláž, A., Madoš, B.: Web Search Engine. In: Acta Electrotechnica et Informatica, Vol. 4, p. 41-45. ISSN 1335-8243, 2011.

[10] Vokorokos, L., Vokorokosová, R.: Pattern matching of partially ordered events in information management systems. In: INES 2008: Proceedings of the 12th International Conference on Intelligent Engineering Systems. February 25-29., 2008, Miami, Florida, p. 169-172. ISBN 978-1-4244-2083-4.

Action graphs and processes of program systems

[1]*Martina ĽAĽOVÁ (4^{th} year)*
Supervisor: [2]Valerie NOVITZKÁ

[1,2]Dept. of Computers and Informatics, FEEI TU of Košice, Slovak Republic

[1]martina.lalova@tuke.sk, [2]valerie.novitzka@tuke.sk

***Abstract*—In this contribution we want to present the processes of program systems by action graphs. We use procesess of architecture SCADA/HMI (Supervisory Control and Data Acquisition / Human Machine Interface). We are focusing on processes in Action Graphs their communication and location of processes. This architecture is considered to be final, deterministic and closed system.**

***Keywords*—processes, action calculus, action graph**

I. Introduction

Action graphs are tools for solving the problem from the perspective of abstraction. They provide a view of system as a complex set of verification features which originate in the design of program system.

Action graphs' approach allows multi-level abstraction. This approach enables monitoring of the behavior on different levels of program system.

System is possible to describe as a formal transcript of behavior our architecture [1],[2], [3]. Proceses of system CADA/HMI are frequently used in industry for Supervisory Control and Data Acquisition of industrial processes and as Human Machine Interface [4].

This is often used for complex architecture with a complex internal structure of processes. The complexity of relations between processes is difficult to describe. We use the action calculus as a tool to describe syntax and action graphs for describe of semantics [5], [6].

II. Program system and action calculus

Program system is possible to describe as a formal transcript of behavior ofprocesses program system, where we use action calculus [1],[2], [3]. This transcript defines interaction between processes and designates specification [7].
Knowing the properties of system and interactions between its components leads to the knowledge of the system. Knowing the composition of the system allows us to understand the structure and its architecture. Let consider understanding of the structure as the understanding of its components and their interactions as partly closed autonomous system.

The program systems in a real environment comprises of two fundamental aspects, data collection and control. In the principle, if we think of control system in real conditions, the most frequent use is in the industry. Industrial control systems are systems which monitor and control industrial processes. Individual industrial control systems are subject to different frameworks, one of which is the SCADA.

SCADA / HMI (Human Machine Interface) provides an interface between man and machine. This interface is structured in a finite number of levels. Each of logical level can be considered as a component of total program system. Each level is organized hierarchically, and allows the investigation of individual components in multiple internal layers.

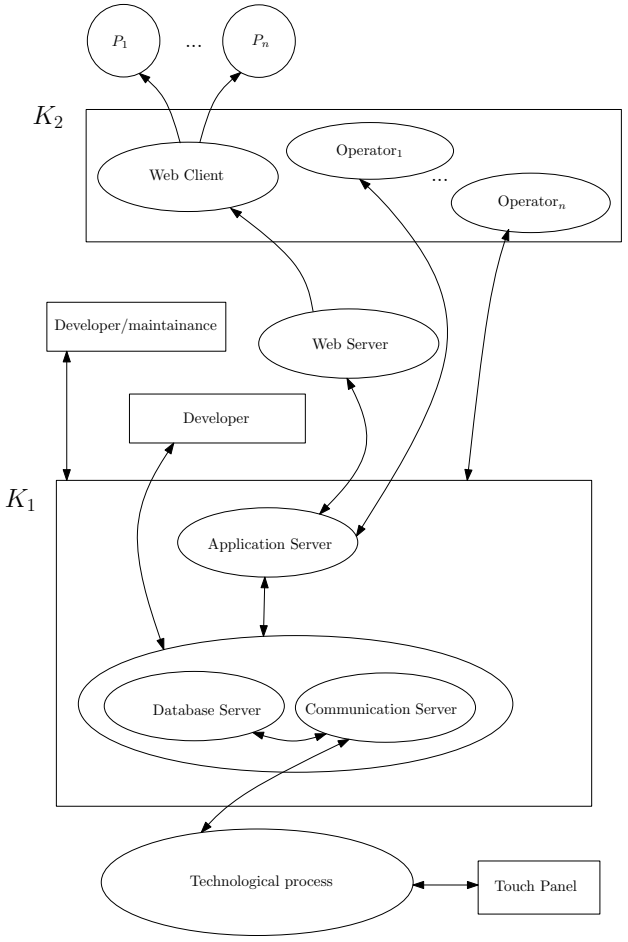


Fig. 1. SCADA/HMI

The action calculus can be defined as a formal method for the description of communicating processes and analysis of its properties.
The process is a series of actions, status changes or functions which lead to results and to fulfillment of events. The event verification of identity and comparision of the access rights we define :

$$a : \{\mathbf{out}(id_P)\}$$

where a is an event of program system, which output event is the identification of the process - id_P.

Through the action calculus we define sublocation α as:

$$\alpha[S : P], \tag{1}$$

where sublocation α is defined by process P in system S
If system S si location and the proces P is a sublocation, this sublocation can be considered as a component of program system. In this architecture by SCADA/HMI framework we can use three components:

- client's component K_2,
- operating/communicating component K_1 and
- technological component K_0.

Program system S by components we can define as:

$$\textbf{def} \;\; K_n \;\; \textbf{in} \;\; S \tag{2}$$

and sef of components

$$K \in \{K_1, K_2, ...\}$$

Component K_2 consists of client's process

$$P \stackrel{\text{def}}{=} x \langle a \rangle .WebClient$$

or processes operator's type and special process $WebClient$ as managing process.
Control processes in action graphs we denoted by pink color, basic activ processes by green and termined processes by black. New process or unused process is denoted by white color.

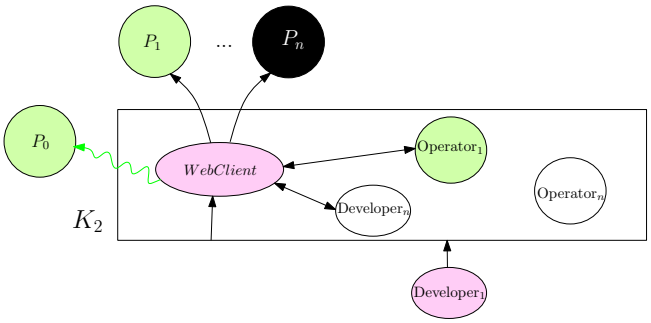


Fig. 2. Action graphs of client's component K_2

Relation between processes of one component are portable by relational rules R on the other component.

$$R \stackrel{\text{def}}{=} K_2 \triangleright P$$

$$R \rightarrow K_1$$

In the component K_1 we define sublocation

$$\alpha[K_1 : ApplicationSever]$$

If we have a sublocation, which is exhausted, we denote this as Ω sublocation

$$\Omega[\gamma : (Db.Server \wedge CommunicationServer)]$$

In this case the sublocation is exhausted, it does not fulfill any events therefore does not perform any actions, probably all processes have been completed.
The processes halt their actions when an event does not longer exist, which have to be executed. The producer of requirements

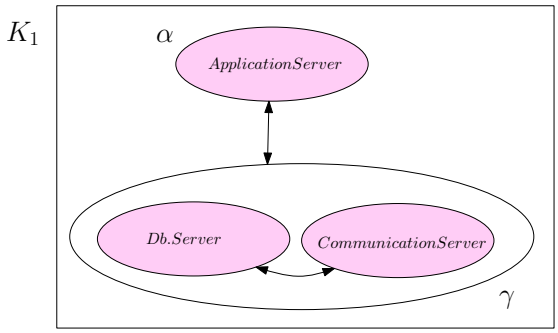


Fig. 3. Action graphs of operating/communicating component K_1

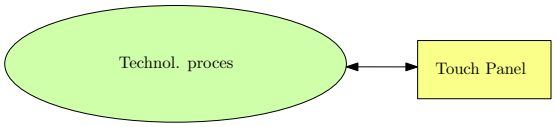


Fig. 4. Technological process in fragment of action graph

for producing events in industrial systems is technological process.
A special process $TouchPanel$ was incorporated into the architecture of program system, because in our view is a black box of behavior, which controls technological process. This process is considered as a component of the simple reason that the panel is directly connected to technological process, but outside of $ComunicationServer$, i.e. outside of the overall management of system. It allows direct communication.
All processes Q_n from technological processes consist of program systems for control purposes of the technological processes.
The technological process is managed and controlled with respect to the source data, that are distributed through communication of program system for all processes. Actual data represent a continuous flow and are recorded in process$Db.Server$.
Relation between processes Q of technological processes and process $Technol.Process$:

$$Q_n \rightarrow Technol.Process$$

we assume, data u in the process $Db.Server$ are from $Technol.Process$:

$$Technol.Process(u) \rightarrow Db.Server.$$

The behavior of the process from the perspective of our program system S of process $CommunicationServer$ by action calculus we write:

$$\begin{aligned} CommunicationServer < a.P \circ a.WebClient \\ \circ\, a.WebServer \\ \circ\, a.ApplicationServer \\ \circ\, a.DbServer > \end{aligned}$$

All program system processes are synchronized. Their communication is running by process $CommunicationServer$. This process is placed in the operating/communicating component K_1. About the synchronization of processes is taken care by special process with actions $tick$.
Let the process *CLOCK* with one event of the set of events U_S and let action $tick$ determine control event e from set of special events U_S, then:

$$e : (tick.CLOCK)$$

Behavior of the process $CLOCK$:

$$CLOCK = (tick \rightarrow CLOCK)$$

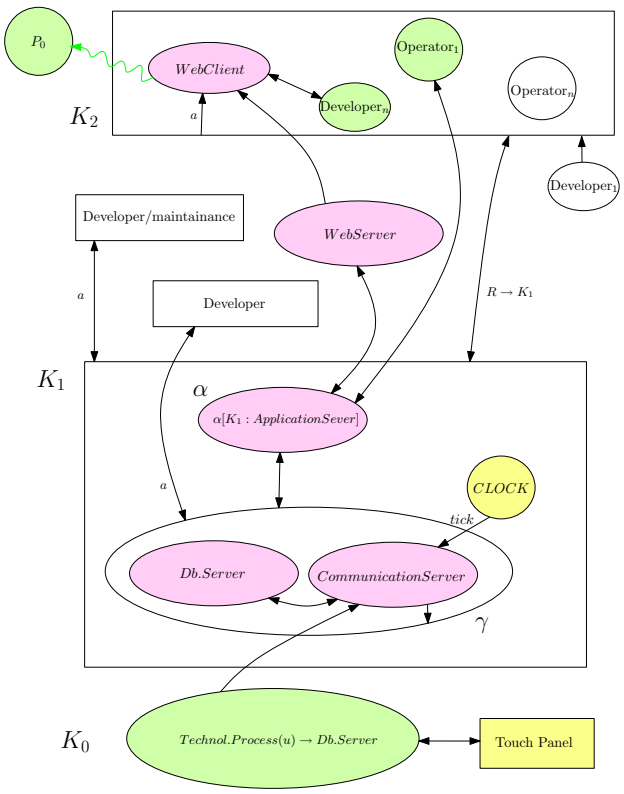


Fig. 5. Behavior in action graph

Class of special processes consits of special processes. Special processes are filling control events by special control actions and they create special events, which form special class of processes. Special processes are fundamental processes accepting in addition to the basic events control actions, too. These special events are fulfilling management control events of program system.

Fundamental processes can use only action in and out for filling basic events.

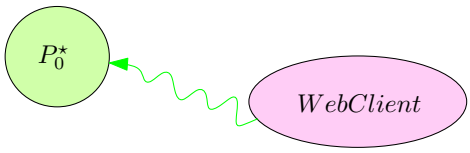


Fig. 6. Mobile proces $P_0^\star$ in action graph, $P_0^\star \rightarrow [K_1 : WebClient]$

Communication in action graph is represented by edge. Edge represents a communication link. This communication link can be mobile. In this case is in the action graph drawn by green corrugated line.

In the component K_2 we can see this case on process P, which is client's process and represents a user. This green corrugated line is mobile connection of user to the process $WebClient$. Process $P^\star$ is a process of base type. Process $WebClient$ is a control process [8], [9].

$$P^\star = \mathbf{out}(a).\mathbf{in}(data).P^\star$$

This way it is possible to observe communication and behavior of program system from mutual processes.

Action calculus provides exact description of events and possibilities of their distribution into the action, which are executed in individual processes.

Any system can be a part of bigger system, or vice versa, may consist of subsystems, wich can be components. It may share common features with other systems, which helps share of information between systems.

Systems are understood as coherent units that behave according to determined rules.

III. Conclusion

Action graphs solve the problem from the perspective of abstraction. They provide a description of a system as a complex set of verification features which originate in the design of control system.

It allows a specific look to individual processes of program system and to the connection between the processes. It allows also to observe the behavior of basic and control processes, thus to verify the behavior of the implemented program system.

Action graphs approach allows multi-level abstraction. This access enables monitoring of the behavior on different levels.

Theory of action graphs provides connection between description of structure and it's processes and relations. The action graphs are used by different action and interaction as a relation between them.

In informatics are action graphs very illustrative description of processes communication.

References

[1] M. Ľaľová and V. Slodičák, "Some useful structures for categorical approach for program behavior," in *Proceedings of CECIIS 2010 Central European Conference on Information and Intelligent Systems, Varaždin, Hrvatska*, 2010.

[2] G. L. Cattani, J. J. Leifer, and R. Milner, "Contexts and embeddings for closed shallow action graphs," Tech. Rep., 2000. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.36.7622&rep=rep1&type=ps

[3] V. Slodičák, "Some useful structures for categorical approach for program behavior," in *Journal of Information and Organizational Sciences, Vol. 35, No. 1*, 2011, pp. 1846–9418. [Online]. Available: www.jos.foi.hr⟨http://www.jos.foi.hr/⟩

[4] S. W. Daneels, A., "What is scada?" In, International Conference on Accelerator and Large Experimental Physics Control Systems, Trieste, Italy, 1999. [Online]. Available: http://accelconf.web.cern.ch/accelconf/ica99/papers/mc1i01.pdf

[5] M. H. Andrew Barber, Philippa Gardner and G. Plotkin, "From action calculi to linear logic." Computing Laboratory, University of Cambridge, Cambridge CB2 3QG, Englan, 2002.

[6] R. Milner, "The space and motion of communicating agents." Lecture Notes for ESSLLI, Department of Mathematics and Statistics, McGill University, 2009.

[7] M. Franeková and comp., "Safety communication of industrial networks." In Slovak EDIS. ŽU Žilina, 2007.

[8] M. Ibnkahla, *Signal Processing for Mobile Communications Handbook*. CRC Press, 2004.

[9] A. D. Gordon and L. Cardelli, "Equational properties of mobile ambients," in *MATHEMATICAL STRUCTURES IN COMPUTER SCIENCE*. Springer, 1999, pp. 212–226.

Analyzing Graphical User Interfaces with DEAL

[1]*Michaela Bačíková (3^{d} year)*
Supervisor: [1]*Jaroslav Porubän*

[1,2]Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

[1]michaela.bacikova@tuke.sk, [2]jaroslav.poruban@tuke.sk

***Abstract*—At present, almost every existing application has a graphical user interface (GUI), which represents the first contact of a user with an application. Therefore the GUI should be created with respect to understandability and domain content. The domain terms, relations and processes should be captured in the GUI so it would be usable. We base our research on this presumption and we created a tool for automated domain analysis, DEAL, which is able to analyze the existing user interfaces and to create a domain model from the gained information. In the current state, we are improving DEAL to be able to derive new relations from the analyzed GUI. In this paper we present the DEAL tool and its features and we also discuss our plans for the future.**

***Keywords*—Domain analysis, Domain extraction, Graphical user interface, DEAL, Component-based software engineering, GUI Domain-specific language**

I. Introduction

Domain analysis (DA) [1] is the process of analyzing, understanding and documenting a particular domain - the terms, relations and processes in the domain. The result of DA is a *domain model* which is used for creation of new software systems. DA is performed by a *domain analyst* [2]. In Table I we summarized three basic sources of domain information currently used to perform DA. The first column lists the particular information sources. The methods for gaining information are listed in the second column and the last column is a summary of disadvantages of the methods for each of the knowledge source type.

According to our research [3] many methods for dealing with the first two information sources exist. The last area, analysis of domain applications, is the least explored area. The main reason is the high level of implementation details – it is hard to extract domain information from code. Reverse engineering deals with extracting relevant information from existing applications, but with the goal of extracting an exact application model, which can be used for implementation of future systems. And since the model is made for further development, it means that whether in an abstract or concrete form, it contains implementation details which again prevent the domain information from being extracted clearly.

Our research deals with extracting information from existing applications, but not from the source code in general. We think that the more appropriate targets for domain analysis are *graphical user interfaces* (GUIs). We stated three hypotheses. First, *users have direct access to GUIs*. Therefore the programmer is forced to use the domain dictionary in GUIs (domain terms and relations between them). Second, a *GUI describes domain processes in a form of event sequences* that can be performed with it. And third, we confirmed that it is possible to *derive relations between the terms* in the UI semi-automatically. Based on our experiments we assume these three hypotheses to be valid when the target application[1] UI is *made of components*.

TABLE I
Different sources of domain information, existing methods for gaining it and disadvantages of the methods

Information source	Method	Disadvantages
Domain experts	interviews questionnaires forms	• no *formalized* form • *time* consuming • depends on *willingness* of experts • requires a *skilled* domain analyst
Domain documentation	artificial intelligence methods natural language processing modeling	• no *formalized* form • *availability* of data sources • *suitability* of data sources – ambiguity of natural language
Domain applications	automatized analysis of • source codes • or databases	• high level of implementation details • no user access ⇒ programmer is not forced to use domain dictionary

Based on these presumptions we designed a method of automatized domain analysis of user interfaces of software applications which we call *DEAL (Domain Extraction ALgorithm)*. The input of the DEAL method is a GUI of an existing application and the output is a *domain model* which can have different forms: a domain dictionary, an ontology, a UML-diagram, etc. A DEAL tool **prototype** which implements this method was created and it will be described in this paper. The DEAL tool prototype supports creating a domain model from an existing UI based on its components. First, it creates a so called *component graph* [4], which is a graphical tree-structured graph of all components located in the target UI. Based on the component graph basic information is extracted from target components and basic relations are derived based on target component types. The result is a *term graph* which represents the domain model. A further research is needed

[1]In the rest of the paper, an existing application which is an input of the domain analysis, we will call the *target application*. A *target component* is a component located in such an application.

to enable deriving more relations. Therefore in this paper we present stereotypes of creating GUIs identified in our experiments with the DEAL prototype. They present a template for designing and implementing more deriving procedures in DEAL.

II. State of the art

Here we briefly summarize the approaches which are used for the domain analysis. The domain models (except for DARE) have to be created manually. Then based on them, different outputs are generated (i.e. software product lines, etc.).

The most widely used approach for DA is the **FODA** (Feature Oriented Domain Analysis) approach [5]. FODA aims for analysis of software product lines by comparing the different and similar features. The **DREAM** approach [6] is based on FODA. The approach is similar to FODA, but with the difference of analyzing domain requirements, not features. Many approaches and tools support the FODA method, e.g. Ami Eddi [7], CaptainFeature [8], RequiLine [9] or ASADAL [10]. Domain model created by FODA is used for further generation of a line of software applications (product line).

There are also approaches that do not only support the process of DA, but also the reusability feature by providing a library of reusable components, frameworks or libraries. Such approaches are for example the early **Prieto-Daz** approach [11] or the later **Sherlock** environment [12].

The latest efforts are in the area of **MDD** (Model Driven Development). The aim of MDD is to shield the complexity of the implementation phase by domain modeling and generative processes. The MDD principle support provides for example the Czarnecki project Feature Plug-in [13], [14] or his newest effort Clafer [15] and a plug-in FeatureIDE [16], [17].

ToolDay (A Tool for Domain Analysis) [18] is a tool that aims to support all the phases of DA.

All these tools and methodologies support the DA process with different features, but the **input data** for DA (i.e. the information about the domain) always come from the users, or it is not specifed where they come from. Only the **DARE** (Domain analysis and reuse environment) tool from Prieto-Daz [19] primarily aims for automatized collection and structuring of information and creating a reusable library by analysing existing source codes and documentation automatically, but not user interfaces specifically.

Very interesting process is also seen in [20] where authors transform **ontology axioms** into application domain rules which is a reverse process compared to ours.

III. The DEAL tool prototype

The DEAL tool prototype is implemented in Java and at present it enables:

- *Loading* the input (a user interface of an existing application).
- *Processing* the input, the output is a component graph.
- *Generating* the domain model, the output is a term graph.
- *Simplifying* the domain model. The simplification in our case means filtering out unnecessary information unrelated to the domain. The output is a simplified term graph.
- *Registering* recorder. The output is the target application with registered recorder of UI events. This enables the user recording events performed in the UI. In later stages we plan to implement replaying these event sequences, which we experimentally proved to be possible.

The DEAL prototype is published as an open-source project on Assembla on https://www.assembla.com/spaces/DEALtool.

A. Domain model representation

We used a graphical tree-like representation of the domain model inspired by the representation used in the Clafer [15] project which is intended for feature modeling in the FODA notation. In DEAL the domain model is perceived rather as an *ontology*. It uses the same notation as FODA, it has however a different semantics. The relations in DEAL do not represent relations between features (functionalities) of an application, rather they represent relations between the *terms* in the given model as it is in ontologies. We used the FODA notation because it is simple and easily readable. Following relation types are supported by DEAL:


Model (MODEL)


Grouping (AND)


Mutually exclusive (MUTUALLY_EXCLUSIVE)


Mutually not exclusive (MUTUALLY_NOT_EXCLUSIVE)

Determining the optional and mandatory items is not relevant in the case of ontology, we therefore use only one type of node labelling (an empty ring), which represents all types of AND groupings. The MODEL relation is a default relation set for the root of the domain model, it has no special semantics. In DEAL tool, each term is represented by an instance of the *Term* class. It captures information such as name, description, icon (sometimes a term in the interface is represented by a graphical icon, which is self-explanatory *and* domain-specific). It contains a link to the target component which it represents. Each node in the graph has a list of its child Term[2] nodes and a relation type, which applies to the children.

B. The DEAL method

The DEAL method contains the following phases:

(1) **Loading and Processing algorithm** - the input is an existing component-based application with a GUI and the output is a *component graph*. In DEAL the input is a Java application and we use reflection and aspect-oriented programming. The Loading and Processing algorithm is invoked for each activated scene[3]. The Loading and Processing algorithm was completely described in [22] as "the DEAL algorithm".

(2) **Generating of a Domain Model** - the input is the component graph generated in (1), the output is a domain model in a non-simplified form (it contains information not related to domain). For each component in the component graph a new Term is generated. If the component does not contain any relevant information (for example a Container), a Term without any domain-relevant information is generated. This is because of preserving the hierarchy of Terms which is stored in the target application by its programmer.

[2]in the rest of this paper, domain terms as real instances (words) will be noted by lower case. Abstract representations of such real terms in the model will be noted by upper case: Term.

[3]We use the term "scene" based on the term by Kösters in [21] A scene represents a window, a dialog or any application frame which contains components.

(3) **Simplification algorithm (SymAlg)** - filtering of information unrelated to the domain, i.e. Terms with no domain-relevant information. The input is a domain model from the generator (2) and the output is a simplified domain model. In current state filtering includes removing multiple nesting and removing void containers. A void container is a node with a component of type Container that does not contain any children.

(4) **Algorithms for deriving relations(DerAlg)** - based on the identification of different types of components, relations between Terms in the model are derived. The input is a simplified domain model; the output is a domain model with relations between Terms. In the current state the DEAL tool prototype is able to derive relations from the different types of components (tab panes, check boxes, radio buttons, lists, combo boxes).

These algorithms are called sequentially in the order they are listed here. SymAlg and DerAlg were defined and implemented based on our previous research in [22].

IV. Analysis of stereotypes of creating GUIs using the DEAL tool prototype

In the current state SymAlg and DerAlg algorithms implemented by the DEAL tool prototype contain only the basic procedures. It is necessary to identify stereotypes of creating GUIs and improve SymAlg and DerAlg based on these stereotypes. In the next chapters we will describe some stereotypes of creating GUIs and based on them we will improve the two algorithms if possible.

We already performed an extensive analysis of stereotypes of creating user interfaces in [22]. After developing DEAL we made a series of experiments and identified additional stereotypes. Here we provide the list of the additional stereotypes and based on them we will design and implement new derivation rules in DEAL.

The experiments were performed with these open source Java applications: Java Scientific Calculator, Java NotePad, Jars Browser, jEdit, Guitar Scale Assistant, iAletU. All of them (except iAlertU [4]) are included into the DEAL prototype on Assembla.

The stereotypes were divided into four groups: graphical, functional, logical and custom components.

A. Graphical

The essential feature of a GUI (if we forget the content) is its graphical representation. If the GUI is made in a haphazard fashion, it hardly provides useful information to its users. The GUI elements should be organized in groupings that make sense.

We identified a number of basic stereotypes of positioning components graphically:

a) **Labels describe other components.**

We call this a *label-for* relationship between a label and a target component. It can be derived for example in Java language by reading the *laberFor* attribute of a JLabel component. The identification with this attribute is implemented in the DEAL tool prototype. In HTML language it is possible to read the *for* attribute.

However the programmers often do not specify the label-for attribute therefore a different approach is needed. Deriving the label-for relationship is possible from the graphical layout of elements. Programmers often put labels horizontally aligned with the target component or above it, aligned left. We discussed the problems of deriving the label-for relationship with our approach in [3].

b) **Separators divide groups of items, which are related to each other but which arent related to items that are separated from this group.**

A separator line (in the Java language represented by the Separator class) is used for example in menus or in forms, where it graphically divides related content from its surroundings. Based on this fact we group the terms representing the components in a separated group into an AND group.

c) **Graphical groupings of components.**

The idea is similar to b). If there is a group of buttons close to each other in a calculator app, and another group of buttons divided from the first group by a blank area, then these two groups should be separated also in the domain model.

B. Functional

The functional components are represented by common control items like buttons, menu items, etc. These are often used to provide functionality, which is common for many applications (like OK, Cancel, Reset, Open, Save, Save As..., New, Exit, Close etc.). If these common terms are discarded from the domain model, then only domain-specific terms will remain.

We have implemented a *"hide as common"* functionality into the DEAL tool prototype, which enables hiding a common used term, so it will be invisible in the domain model. These hidden terms will be added to a database that will contain a collection of the most common terms. Based on this database the tool will suggest to hide these terms also in other models. This will ensure the tool will learn.

C. Logical

From some components relations can be logically derived. For example a group of terms representing radio buttons will be mutually exclusive. A group of terms representing check boxes will be not mutually exclusive, but related to each other. The same is true for combo boxes (mutually exclusive) and lists (with a single selection - mutually exclusive, with multiple selection - mutually not exclusive), tab panes (mutually exclusive) etc.

Deriving of these relations is implemented in the DEAL tool prototype.

D. Custom components

Custom components are components programmed by a programmer. These components can serve for gaining domain-specific information.

For example during the experiment with the open-source Launch4j application we encountered the **TitledSeparator** component. It is a normal content separator but it also contains a title. This can be used to describe the group of separated terms. In DEAL new handlers for custom components can be implemented by simply extending the **DomainIdentifiable**$<T>$ abstract class, defining the component class as T and by implementing abstract methods

[4] the open source jar for this application does not exist anymore

for gaining domain information. The handler should also be registered in the list of implemented handlers **domainIdentifiables.properties**. Then the DEAL tool prototype will take also this special type of component into account.

V. Conclusion

In this paper we introduced our DEAL method for analysing existing software applications for domain terms and for extracting a domain model in a form of a term graph.

We introduced and described our DEAL tool prototype which enables creating domain model from a GUI of an existing Java application. For the future implementation of simplification procedures is planned and implementation of a number of procedures for deriving new relations based on the stereotypes of creating graphical user interfaces which were identified in this paper.

We also plan to implement the functionality for saving the domain model into one of the ontological formats (e.g. OWL) and XML. DEAL supports a recording feature which enables to record the event sequences made by the UI user into a Term sequence. In the final phase we plan to implement a replay feature to replay this sequence on the target application. This replay feature could be utilized as macro executor, but more importantly it's utilization lies in educative processes. If one person in one company knows how to perform a task on his/her application, he or she could help their friend in another company (or town) by sending a record of tasks performed on the application. The friend could then download this record and use the replay function on an actual application. Unlike user guides or tutorials, which are rather static, the replay function works on a real application and in real-time, therefore it is more understandable and visual.

Our method and tool can serve as an additional process of domain analysis, where the domain analyst does not start from scratch when creating a domain model, but he can gain a simple domain model from existing application. We plan to apply our research in the field of usability evaluation – specifically domain usability evaluation where it can be utilized in two areas: i) evaluation of domain dictionary of an existing application if it matches the real world; and ii) evaluation of UI event sequences if they are correct and match the domain processes in the real world.

Acknowledgment

This work was supported by VEGA Grant No. 1/0305/11 Co-evolution of the artifacts written in domain-specific languages driven by language evolution.

References

[1] J. Neighbors, "Software construction using components," Ph.D. dissertation, University of California, Irvine, 1980.

[2] ——, *The Draco approach to constructing software from reusable components*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1986, pp. 525–535.

[3] M. Bačíková and J. Porubän, "Defining computer languages via user interfaces," Master's thesis, Technical university in Koice, Faculty of Electrotechnical Engineering and Informatics, 2010.

[4] ——, "Automating user actions on gui: Defining a gui domain-specific language," in *CSE 2010: proceedings of International Scientific conference on Computer Science and Engineering*, 2010, pp. 60–67.

[5] K. C. Kang, S. G. Cohen, J. A. Hess, W. E. Novak, and A. S. Peterson, "Feature-oriented domain analysis (foda) feasibility study," Carnegie-Mellon University Software Engineering Institute, Tech. Rep., November 1990.

[6] M. Moon, K. Yeom, and H. Seok Chae, "An approach to developing domain requirements as a core asset based on commonality and variability analysis in a product line," *IEEE Trans. Softw. Eng.*, vol. 31, pp. 551–569, July 2005.

[7] K. Czarnecki, T. Bednasch, P. Unger, and U. W. Eisenecker, "Generative programming for embedded software: An industrial experience report," in *Proceedings of the 1st ACM SIGPLAN/SIGSOFT conference on Generative Programming and Component Engineering*, ser. GPCE '02. London, UK: Springer-Verlag, 2002, pp. 156–172.

[8] "Captainfeature, the webpage of captainfeature sourceforge.net project," https://sourceforge.net/projects/captainfeature, 2005, [Online 2011].

[9] "The webpage of requiline project," https://www-lufgi3-informatik.rwth-aachen.de/TOOLS/requiline/index.php, 2005, [Online 2011].

[10] P. S. E. Laboratory, "A review of asadal case tool."

[11] R. P. Díaz, "Reuse Library Process Model. Final Report," Electronic Systems Division, Air Force Command, USAF, Hanscomb AFB, MA, Technical Report Start Reuse Library Program, 1991.

[12] A. Valerio, G. Succi, and M. Fenaroli, "Domain analysis and framework-based software development," *SIGAPP Appl. Comput. Rev.*, vol. 5, pp. 4–15, September 1997.

[13] K. Czarnecki, M. Antkiewicz, C. Kim, S. Lau, and K. Pietroszek, "fmp and fmp2rsm: eclipse plug-ins for modeling features using model templates," in *Companion to the 20th annual ACM SIGPLAN conference on Object-oriented programming, systems, languages, and applications*, ser. OOPSLA '05. New York, NY, USA: ACM, 2005, pp. 200–201.

[14] M. Antkiewicz and K. Czarnecki, "Featureplugin: feature modeling plug-in for eclipse," in *Proceedings of the 2004 OOPSLA workshop on eclipse technology eXchange*, ser. eclipse '04. New York, NY, USA: ACM, 2004, pp. 67–72.

[15] K. Bak, K. Czarnecki, and A. Wasowski, "Feature and meta-models in clafer: mixed, specialized, and coupled," in *Proceedings of the Third international conference on Software language engineering*, ser. SLE'10. Berlin, Heidelberg: Springer-Verlag, 2011, pp. 102–122.

[16] T. Thum, C. Kastner, S. Erdweg, and N. Siegmund, "Abstract Features in Feature Modeling," in *Software Product Line Conference (SPLC), 2011 15th International*. IEEE, Aug. 2011, pp. 191–200. [Online]. Available: http://dx.doi.org/10.1109/SPLC.2011.53

[17] T. Thum, D. Batory, and C. Kastner, "Reasoning about edits to feature models," in *Proceedings of the 31st International Conference on Software Engineering*, ser. ICSE '09. Washington, DC, USA: IEEE Computer Society, 2009, pp. 254–264.

[18] L. Lisboa, V. Garcia, E. de Almeida, and S. Meira, "Toolday: a tool for domain analysis," *International Journal on Software Tools for Technology Transfer (STTT)*, vol. 13, pp. 337–353, 2011.

[19] W. Frakes, R. Prieto-Diaz, and C. Fox, "Dare: Domain analysis and reuse environment," *Ann. Softw. Eng.*, vol. 5, pp. 125–141, January 1998.

[20] O. Vasilecas, D. Kalibatiene, and G. Guizzardi, "Towards a formal method for the transformation of ontology axioms to application domain rules," *Information Technology and Control*, vol. 38, no. 4, pp. 271–282, 2009.

[21] G. Kösters, H.-W. Six, and J. Voss, "Combined analysis of user interface and domain requirements," in *Proceedings of the 2nd International Conference on Requirements Engineering (ICRE '96)*, ser. ICRE '96. Washington, DC, USA: IEEE Computer Society, 1996, pp. 199–. [Online]. Available: http://dl.acm.org/citation.cfm?id=850944.853112

[22] M. Bakov and J. Porubn, "Analyzing stereotypes of creating graphical user interfaces," *Central European Journal of Computer Science*, vol. 2, pp. 300–315, 2012. [Online]. Available: http://dx.doi.org/10.2478/s13537-012-0020-x

Application Updater Based on OSGi Platform

[1]*Marek Novák (4^{th} year),* [2]*Martin Kiss (2^{nd} year master)*
Supervisor: [3]*František Jakab*

[1,2,3]Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

[1]marek.novak@tuke.sk, [2]martin.kiss@student.tuke.sk

***Abstract*—Although the Java platform is one of the mostly used development technologies, it does not provide sufficient support for automated update of an application. Frequently used approach for this purpose is JNLP, however it suffers of many flaws. We have utilized an OSGi platform to create a framework for configurable automated updater, where our framework is a set of pluggable bundles. The framework allows to configure frequency of an update and various sources (protocols) as: http, https, git, svn. Configuration may be realized in common OSGi configuration file or separately in a manifest file of each bundle.**

***Keywords*—bundle, classloader, git, OSGi, svn, update**

I. Introduction

At first look, implementation of an automatic update mechanism seems to be a trivial task, already supported in many applications and implemented by various projects. All of us are already familiar with a bothering system tray icon in Windows operating system trying to notify a user of a new Java update, or an operating system shutdown process got stuck with downloading new patches and very well known message *"Please do not turn off or unplug your machine . . . "* Naturally, it occurs at a time you are in a hurry for a tram or your partner is waiting for you.

Although automatic update of an application is not new, there are not available many frameworks providing this functionality. Even when we are speaking about highly spread Java platform, to which this paper is devoted. When a developer starts to search for a suitable framework, he quickly gets to *Java Web Start* framework and *Java Network Launch Protocol* (JNLP) protocol [1]. This technology is frequently used, however suffers with many flaws and has an unpredictable behavior. Problems occur when installing multiple versions or the same version multiple times, which is a common scenario in development phase or when deploying to Quality Assurance (QA) environment. In our opinion, OSGi platform and corresponding frameworks as *Apache Felix* or *Eclipse Equinox* are highly suitable for an application that requires to be automatically updated. OSGi natively supports changing a source code of an application while running, so the application may change itself.

In this paper we present a configurable framework based on OSGi platform we have developed, that automatically updates an application from various sources as http, https, git and svn.

II. Current solutions

A. Java Network Launch Protocol (JNLP)

JNLP is a protocol defined in XML format that specifies how to start an application, the source of a jar package and other rules defining how exactly the launching mechanism would behave. The implementation of JNLP is *Java Web Start* (JWS) framework. JWS allows to launch Java applications using a web browser, however unlike Java applets, JWS applications do not run inside the browser, but they run as regular desktop applications. The main advantages of JWS deployment as outlined in [2] are: support for multiple platforms, where JWS client will download and use correct native libraries; automatically downloaded and installed JRE if not presented; sandboxing model; and automatic updates, where JWS client can check the server if there is a new version of the application available and automatically download and install it.

Although generally useful and widely used JWS technology has its limitations and flaws. Some of them are [3]:

- Complex interactions of web caching proxies, browser cache and JWS are very difficult to debug and sometimes only complete reinstallation of the applications and JWS resolves the situation.
- JNLP is unnecessarily complex, which makes it fragile. Also includes many features which are often not used.
- Automatic JVM downloading often doesn't work.

B. Background Update Process

Many applications implement updater as a separate background process, that checks in a predefined time interval for a new version. If a new version of an application is presented, they shutdown the application, download new version and restart. One of many examples is Java Auto Update [4] that downloads new versions of JRE. Among the disadvantages of this solution is the need to implement the updater from a scratch and a separate process itself, so for one application there are installed actually two.

C. Workaround With a System Script File

To have only one application that can update and restart itself, the program may be implemented by the usage of a command line interpreter script file. In this scenario a Java application downloads a new jar file from a remote resource, runs the script and shutdowns itself. The script waits a predefined time interval until the application is stopped and runs a new downloaded version. Besides the fact, that this solution is "ugly" for any programmer, it is vulnerable, because the application does not know anything about the state of a script file, nor the script file has any knowledge about the application.

D. OSGi Bundle Repository (OBR)

OBR is a service directly defined in OSGi specification providing access to a set of bundle repositories [5] [6]. It simplifies the process of deployment, takes care of deployment dependencies and provides updating mechanism as well. This service is already supported by majority of OSGi frameworks and can be used for automated update of bundles, however only in case the bundle is part of a public repository. The service is oriented rather on communities and was not aimed as a single solution. Communities like *Apache Felix*, *SpringSource* or *Apache Sling* have their own repositories, but they are privately maintained. Although it is possible to create a private repository by e.g. OBR Apache bundle, we think it is not suitable for a simple updating mechanism, since it requires a separate server and adds needless overhead.

E. Our Novelty

Our aim was to provide a simple framework with a small footprint offering fast integration with OSGi. As we have already stated, the auto updating mechanism is a standard feature solved many times in a lot of applications, nonetheless still missing in OSGi environment. There are accessible two OSGi frameworks with provisioning functionality: *Apache Ace* [7] and *Equinox P2* [8]. The idea of both is to a provide a server that is responsible for managing requests of many update-greedy clients. For instance, an *Eclipse IDE* is built on OSGi and uses *Equinox P2* to provide a centralized server for updating plugins.

In contrary, our idea is to have self contained OSGi instances without the need of any plugin server installation. In our focus is a scenario of a developer who is hosting his project on *GitHub* or *Bitbucket* demanding to add quickly and seamlessly auto-update support to his project. Naturally, if he uses Java technologies (or languages compiling to bytecode).

Closely related is a project *Getdown* [3] aimed to substitute JWS by providing a functionality to download a collection of files and upgrading those files if needed. However, it does not support application restart, update scheduling, etc. It just download a new version of a file if provided, nothing more. Auto-restart must be implemented externally, e.g. by an *ANT* task.

From the above we can sum up the intended features of our framework to following points:

- various sources: HTTP, Git, SVN
- automatic application update and restart directly from the same process (JVM is never shut down)
- update scheduler
- no server installation required (like *OBR*, *Apache Ace*, *Ecquinox P2*)

III. OSGi Platform

The OSGi technology brings modularity in Java programming language allowing applications to be constructed from small, reusable and collaborative components. Each component, called *"bundle"*, has its own life cycle and may be remotely installed, started, stopped, updated and uninstalled without requiring a reboot of the system [9]. The bundles must define only dependencies between them through interfaces; the searching and binding of required services is handled by OSGi framework. There are available several OSGi implementations as *Knoplerfish*, *Apache Felix*, *Equinox* and *Concierge OSGi*.

Although the OSGi provides many benefits, from a view of automated updater there are important two of them: source code reload during runtime and native versioning support.

A. Versioning

Versioning is natively supported by OSGi. A bundle has its own version and can specify a version for each package being exported as well as version ranges when importing packages. Version is build from four parts: major, minor, micro, and qualifier [9]. Version number does not need to be specified in a JAR filename, but must be defined as a manifest header, e.g.: *Bundle-Version: 1.0.0.1*.

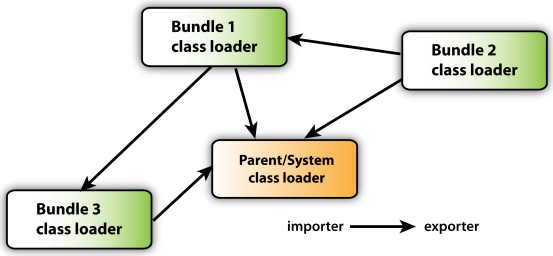


Fig. 1. OSGi class loader delegation model

B. Class Loader

The key feature of OSGi is its sophisticated class loading mechanism. In contrast with the standard JVM class loaders structure which is tree based, class loaders in OSGi create a graph (Fig. 1). Each bundle has its own dedicated class loader and based on export/import declarations in manifest files, the bundle loads a class from another bundle's class loader. The mechanism of loading classes in OSGi is more complex [10], however for the purpose of our automated updater we do not go into deeper details.

What is important, when a bundle is uninstalled and there are no references to exported classes, the bundle's class loader is garbage collected. Our updater utilizes this behavior. When there is available new version of a bundle, the framework stops and uninstalls the old version, downloads new one and starts it. The class loader from the old bundle is garbage collected and therefore old source code is discarded and new source code is loaded by a new bundle class loader (Fig. 2). All this happens while the OSGi instance is still running, therefore restarting the application is not needed.

IV. Updater Architecture

Our framework, with a working name *"oUpdater"*, is built as a set of three bundles. The core bundle reads the configuration, has access to all bundles of the environment through *BundleContext* object, updates other bundles, schedules the updater timer, and supports *http* and *https* protocols for downloading new versions of bundles (Fig. 3).

Having in mind component oriented programming principles of OSGi specification and our aim to create a framework with a small footprint, we detached the support of version control systems *git* and *svn* into separate bundles. Bundle for git uses an external library *JGit* [11] and bundle for svn uses *SVNKit* [12]. The both propagate its functionality by registered OSGi service.

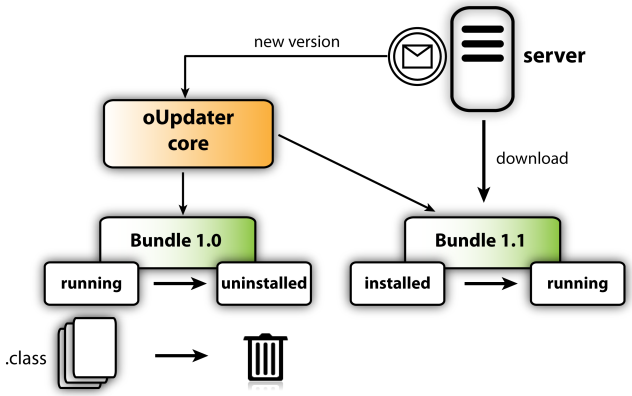


Fig. 2. Updating process from Bundle version 1.0 to version 1.1

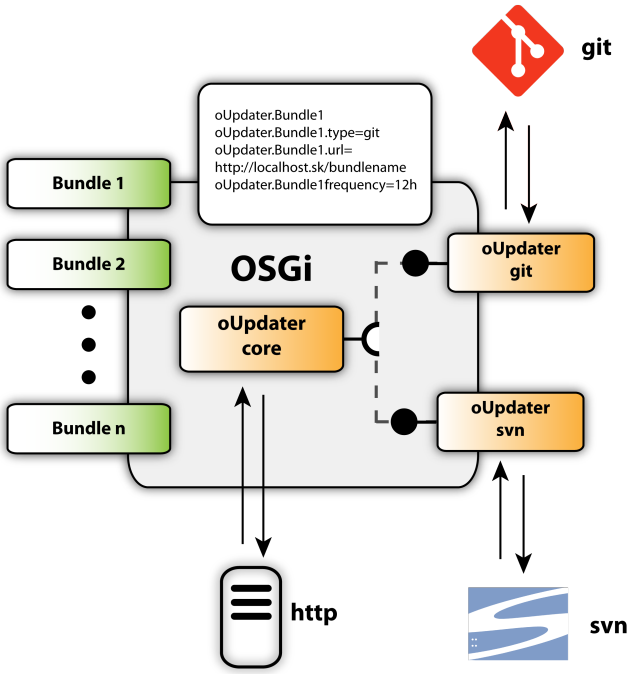


Fig. 3. Framework oUpdater in OSGi context

V. Updater Configuration

A. Configuration Files

To avoid unwanted complexity and to make the configuration as easy as possible we stick with standard configuration files used in OSGi:

- framework common configuration file
- manifest file

Framework configuration file is in a format of Java properties file, frequently used to store parameters in Java related technologies. In OSGi framework it is used for common framework parameters like bundle storage location, cache location, starting level and also external bundle's parameters as e.g. http server port. The properties file is only a key value store and therefore is not hierarchical. However, the hierarchy may be supported by some special characters, usually a dot. In our case three levels are absolutely sufficient. The first level identifies an updater, the second a bundle that should be configured and the third a particular parameters. MANIFEST.MF file is a common part of java archive (JAR) or web application archive (WAR). OSGi framework adds additional headers as bundle name; imported and exported packages; activator class; etc. We add our own manifest headers.

B. Configuration Structure

The example of configuration in common properties file is as follows:

```
oUpdater.BundleName
oUpdater.BundleName.type=git
oUpdater.BundleName.url=http://ex.sk/bundlename
oUpdater.BundleName.path=dist/bundle-1.0.0.1.jar
oUpdater.BundleName.frequency=12h
oUpdater.BundleName.time=00:50:00
```

Since properties file is not structural, we use a dot sign to build a hierarchy. Our configuration consist of three levels, where in `oUpdate.BundleName.type` `oUpdate` is the first level, `BundleName` the second and `type` is the third. Thus we can configure as many bundles as needed in one configuration properties file.

First level `oUpdate` only specifies that this configuration belongs to our framework. Felix common configuration file *conf.properties* contains the configuration for the whole OSGi instance, therefore we wanted to enhance readability.

Second level specifies name of a bundle as defined in manifest file of a particular bundle with header *Bundle-SymbolicName*. The bundle symbolic name uniquely identifies a bundle. Because the symbolic name often contains a dot sign it must be declared on a separate line to allow correct parsing of further parameters.

Configuration in manifest file is very similar, just notation differs a little. Bundle name is not required, because manifest file is deployed with and belongs to only one bundle. Parameter `oUpdate.BundleName.type` in manifest file is `oUpdater-Type`, the others are built likewise.

C. Parameters

Parameters on a third level are as follows:

- **type**: Defines a protocol, how the new version of a bundle is accessible. Currently it may acquire following parameters: *http*, *https*, *git* and *svn*.
- **url**: Defines new version resource URL. For *git* it is e.g. *https://jankohrasko@bitbucket.org/myProject/myBunle.git*.
- **path**: This is an optional parameter and defines a path in cloned git repository to the bundle. By default a framework searches for a jar file with name *SymbolicName-x.x.x.x.jar* in a root folder of a git or svn repository. If not found this parameter is required.
- **frequency**: May be defined in days with suffix *'d'*, hours with suffix *'h'* or minutes with suffix *'m'*. So for instance value '12h' means that updater will look for a new version every twelve hours.
- **time**: Specifies a precise time when the updater should look for a new version of a bundle. This is an optional parameter and may be used only if frequency is defined in days.

VI. Conclusion

Implementation of an automated updater in Java is possible through several approaches as background updater process, JNLP, etc., however all of them suffer of some drawbacks. OSGi is steadily getting more attention and is already spreading in enterprise environment. We have utilized this technology

to implement our own framework for an automated configurable updater *oUpdater*. The framework is capable to use more sources as http, git, svn and thus supporting popular repositories like *GitHub* or *Bitbucket*. Accordingly it schedules the update for each bundle separately. The benefits are that application is able to update itself without the need of the restart and is easily configurable.

ACKNOWLEDGMENT

This work is the result of the Project implementation: Competency Centre for Knowledge technologies applied in Innovation of Production Systems in Industry and Services, ITMS: 26220220155, supported by the Research & Development Operational Programme funded by the ERDF.

REFERENCES

[1] J. Zukowski, "Deploying software with jnlp and java web start," Aug. 2002, accessed: March 2013. [Online]. Available: http://www.oracle.com/technetwork/articles/javase/index-135962.html
[2] "The java tutorials: Lesson: Java web start," accessed: March 2013. [Online]. Available: http://docs.oracle.com/javase/tutorial/deployment/webstart/index.html
[3] M. Bayne, "getdown: Major design goals," Dec. 2012, accessed: March 2013. [Online]. Available: https://code.google.com/p/getdown/wiki/Rationale
[4] J. H. Page, "What is java auto update? how do i change notify settings?" 2012, accessed: March 2013. [Online]. Available: http://www.java.com/en/download/help/java_update.xml
[5] W. Gédéon, *Osgi and Apache Felix 3.0 Beginner's Guide*. Packt Publishing, Limited, 2010.
[6] A. F. page, "Apache felix osgi bundle repository (obr)," 2009, accessed: March 2013. [Online]. Available: http://felix.apache.org/site/apache-felix-osgi-bundle-repository.html
[7] "Project page: Apache ace," 2013, accessed: April 2013. [Online]. Available: http://ace.apache.org/
[8] "Project page: Ecquinox p2," 2013, accessed: April 2013. [Online]. Available: http://wiki.eclipse.org/Equinox_p2
[9] A. De Castro Alves, *Osgi in Depth*, ser. Manning Pubs Co Series. Manning, 2011.
[10] O. Alliance, "OSGi Service Platform, Core Specification, Release 4, Version 4.2," OSGI Alliance, Tech. Rep., Sep. 2009.
[11] "Jgit home page," accessed: March 2013. [Online]. Available: http://eclipse.org/jgit/
[12] "Svnkit home page: [sub]versioinng for java," accessed: March 2013. [Online]. Available: http://svnkit.com/

Attribute-Oriented Programming and Its Challenges: Overview

[1]*Milan NOSÁĽ (2^{nd} year)*
Supervisor: [2]Jaroslav PORUBÄN

[1,2]Department of Computers and Informatics, FEEI TU of Košice, Slovak Republic

[1]milan.nosal@tuke.sk

***Abstract*—This paper introduces attribute-oriented programming, an extension of programming paradigms that allows programmers to annotate program elements with custom metadata. Although this technique exists for some time, there are still many related challenges that remain untackled. This overview presents selection of them, the ones that we consider the most important, and outlines and discusses possible directions to solve them.**

***Keywords*—attribute-oriented programming, annotations, attributes, programming languages**

I. Introduction

Attribute-oriented programming (@OP) is a technique that is used nowadays for many different purposes. Mainstream usages include code generation ([18], [13]) and configuration ([14], [17]). In the industry there are many frameworks that started using Java annotations as a configuration format as an alternative to long-used XML documents. We can mention the most famous as Spring framework [16], Hibernate [9], EJB (Enterprise Java Beans) [6], JAXB (Java Architecture for XML Binding) [7], JAX-WS (Java API for XML Web Services) [8] or JPA (Java Persistence API) [6].

After more than 10 years of its existence one would expect that this technique would be well explored. But if we look closer, we will find many loose ends in the @OP basis. Obviously, 10 years was not enough for @OP to mature. The goal of this paper is to present and discuss some of the most important challenges in the field of the @OP.

II. Attribute-oriented programming

Attribute-oriented programming (@OP) has been around for some time already. We can find some of its ideas behind attribute grammars or literate programming. Probably the first true implementation of attribute-oriented programming can be considered XDoclet, tool that was mainly used in EJB to generate necessary code for the EJB architecture from attributes [4]. Microsoft recognized the importance of the @OP and included its implementation to .NET platform from the very beginning (2001) as attributes. Soon after that (2004) Sun introduced Java annotations with Java 1.5.

Attribute-oriented programming is an extension of classic programming paradigms like object-oriented programming, procedural programming or functional programming. Attribute-oriented programming extends syntax of programming languages by adding rules that allows tagging (annotating) program elements (such as class, function, etc.) with declarative structured metadata. This makes attribute-oriented programming rather an extension than a new paradigm [2]. Wada [19] defines @OP as a program-level marking technique, thus implying the difference between the programming paradigm and attribute-oriented programming.

This is the reason why we can also meet with a term attribute-enabled programming (attribute-enabled software development coined by Vasian Cepa in [2]), that tries to distinguish @OP from programming paradigms. This term better reflects nature of attribute-oriented programming, but today community is already used to term attribute-oriented programming. This is the reason why I will use the term @OP too.

The main idea behind @OP is to embed metadata directly in the source code. Embedding metadata in the code is the main difference from other metadata facilities that brought @OP very quickly to the top. Of course it did not substitute external metadata, but it came as a great alternative. Comparison of internal and external metadata (XML and @OP) can be found in [14], [17].

Structured metadata that annotates program elements are called annotations or attributes. As Cepa [2] states, these annotations are usually used to model domain-specific abstractions (DSA). DSA are abstractions that are specific for a given domain, for example if we consider domain of data persistence in relational database, we can speak of a table DSA, or a column DSA. Annotations provide a technique to easily implement embedded domain-specific language [10] for some software product line [5] to express these DSA directly in the code. This approach can boldly reduce repeating code needed for implementation of domain-specific concerns that can be abstracted using DSA. More about this aspect of @OP can be found in [2].

III. Challenges

In this paper we do not want point out all the issues that can be found in the field of the @OP. We believe we don't even recognize most of them, because we consider research in this field merely beginning its journey. Our goal is just to discuss few of them we consider most important and up-to-date. Let us have a look at our choice.

A. @OP alternatives

Attribute-oriented programming is a technique of marking program elements [19]. The main point of this technique is to embed some custom metainformation directly to the source code. Of course @OP is not the only way how to embed metadata into the code. We think that exploring options

in alternative approaches to embedding metainformation into code is an important challenge, although more theoretical than practical.

Cepa in [2] explores and compares two different approaches from the aspect of keeping the architectural decisions made in model (UML) in source code[1]. He discusses pseudosyntactic rules and interfaces.

Pseudosyntactic rules are conventions made about writing code for some domain-specific purpose. Probably the best known pseudosyntactic rule is writing getters and setters in Java beans. Getter should start with prefix "get" following by the name of the accessed field with first letter uppercase. This convention is not enforced by compiler. Therefor we call it pseudosyntactic – its syntax is not enforced by the compiler's syntactic analyzer. But breaking these pseudosyntactic rules in better case causes runtime error, in worse silent (without warnings or errors) unwanted behavior.

Getters and setters are just an example of pseudosyntactic rules. Using these conventions one can express probably anything that can be expressed by annotations[2]. But using annotations gives it more structure and what is more important, it is syntactically checked. Not to mention the support from language tools like code completion in IDE.

Or one can use marking interfaces discussed by Cepa [2]. Implementing a marking interface carries the same expressive information as annotating the class with an annotation without properties. We can also use interfaces with methods as an alternative to annotations with properties, where properties are simulated by implementing the methods of the interface. And these are merely two of the alternative approaches to embed metadata into code.

B. Annotations' usage restrictions

Annotations' usage restrictions are one of the most recognized challenges. Three different tools have been developed by Cepa et al. [3], Noguera et al. [11] and Ruska et al. [15]. All these authors recognized a need to formally specify dependencies between annotations and to have a tool (preferably compiler) that would check them during compilation. This way a programmer can get warnings and errors before deployment and he or she can avoid problems cause by runtime errors or unwanted behavior.

ADC tool (Attribute Dependency Checker) by Cepa et al. [3] is the beginning of the discussion of attribute dependencies (ADC works on C# .NET attributes). ADC presents simple validation tool that supports some types of dependencies between attributes. AVal by Noguera et al. [11] came on the scene soon after that to deal with the issue on the Java platform. AVal is more complex and covers more relations between annotations. The tool introduced by Ruska et al. [15] is also dedicated for the Java annotations and although it seems a little less strong than AVal, the work by Ruska coins a very important term in this context – annotation constraint patterns. Annotation constraint pattern is a type (a kind) of a relationship or a constraint that applies on the annotation usage. The term constraint instead of the dependency is crucial, because sometimes the pattern defines the applicability on the source code without dependency on any other annotation.

We see a potential for research in this field in exploring and discussing annotation constraint patterns. Existing tools and works provides great base on which the research can be built. We also hope that more solid work would induce standards in this matter, so the validation will become a language vendor's responsibility.

C. GAAST-languages

GAAST-language (Generalized and Annotated Abstract Syntax Tree) is a language that supports generalized and annotated abstract syntax tree (GAAST) API. GAAST is an generalized abstract syntax tree that holds appended attributes (annotations) at its nodes (concept of GAAST was introduced by Cepa [2]). Generalized in this context means the tree is generalized for three main stages of program life cycle - compile time, load time and runtime.

This abstract syntax tree (AST) is traversed to read information about program elements. AST itself can be considered metadata about program elements, but this metadata is inherent from programming language. Appended annotations can carry additional metadata that is purposely added by a programmer to express some custom information. This custom information is used to generate code or as a configuration for frameworks to introduce DSA into programs. Thus the GAAST should be the main interface for DSA programmers to access information about DSA instantiations[3].

So far we cannot talk about full GAAST-language in context of current programming languages. Although both two mainstream object-oriented (OO) languages (C# and Java from version 1.5) support attributes (annotations), they support merely strict APIs for manipulating them and no generalized AST. Cepa et al. [2] worked out an idea how to support kind of a full GAAST using existing technologies in C#, but that is merely his work and not standard supported by Microsoft at all.

Supporting a full GAAST in a language by its vendors makes implementation of DSA using annotations very effective and flexible [2]. Evolution of software product line based on @OP is less fragile (evolving language does not harm GAAST implementation, because it is also evolved by language vendors). Therefor we believe that there is an urgent need to explore options in standardizing a GAAST for programming languages – there are probably some rules that could be used to find GAAST for any language, independent of programming paradigm. We believe this idea is a challenge worth exploring.

D. Annotating unnamed program elements

This issue is closely coupled with the issue of the GAAST-languages. Although there is a logic in this matter, annotations in Java and attributes in C# can be used merely on named program elements (such as class, method, field, variable, etc.). One cannot annotate a loop or a block. This could be caused

[1] Although this is just one aspect and not general comparison, many of the conclusions may be generalized. This work is important for the beginning research in this field.

[2] Annotations with properties (or attributes) have the same expressive strength as annotations without properties (we are not going to prove it, proof can be found in [2]). Instead of using annotation without property we can use the name of the annotation (the name carries the important information) in the name of the class, method or field as a pseudosyntactic rule. This simple mapping shows the equality of the two approaches.

[3] DSA definition is a specification of some domain-specific concern that can be abstracted, DSA instantiation is a concrete instance that is using feature supported by the DSA

by implementation issues, but we can also see this as a step towards GAAST. Loops, blocks and statements do not exist in runtime anymore – they are lost during the compilation[4]. Of course annotations on these program elements are accessible during the compile-time on the source code level, but if we will not preserve full AST of the language in the compiled program, it cannot be processed in the runtime (or even loading time).

Preserving whole AST in the bytecode seems too demanding and difficult. We can solve things allowing annotating unnamed program elements with annotations with source retention time, explicitly stating that they will be lost during compilation, but this would be in conflict with the GAAST idea. The question is, whether this drawback is worth it. In our work about the elucidative programming in Java [12] we are discussing possibility of using annotations to mark unnamed code blocks so they could be identified by a name (in fact using annotations on unnamed program elements to make them named). Although this is just one case where annotations on unnamed program elements could be helpful, there is a space for research to find reasons for support of annotating unnamed program elements.

E. Moving to other paradigms

One can hardly miss the direction in the industry and in the research as well in @OP implementation on programming paradigms. The most ado about @OP is in the object-oriented paradigm (most of the sources that can be found on the internet concerns @OP in Java or .NET). Therefore we find interesting looking at the possibilities in implementing and using annotations in other paradigms (functional programming, etc.) as well. Interesting could be also trying to mix paradigms (their characteristics) using annotations. Mixed paradigms are quite common nowadays, we can mention lambda functions in C++11 [1]. But most of the time these require changing the language grammar. Annotations as an extensibility point of the languages can provide a tool for less invasive introduction of such elements into the language.

F. When and how to use annotations

One of the key questions in the field of the @OP is when and how to use annotations. Almost anything in the source code can be expressed through the annotations. On the internet there are many discussions on the blogs and forums about the usage of the annotations [20]. These discussions are more or less about annotations themselves, about using them for different purposes (like configuration of multiple frameworks). This is a clear sign that this question should be elaborated and the answer would be welcome.

Cepa states that annotations should be used only when there is not a better way to reach the goal [2]. We don't want to question this lemma, it is truth indeed, but if we think about it, it is too general. Of course we want to use annotation merely in situations where there is not a better way to get the result. We want to come with more specific answers, with a rules or recommendations that lead to good habits in using of the annotations. Of course this might not be possible, but it is surely worth exploring. Answering the question could be a success as the design patterns in object-oriented programming.

[4]There is also one interesting exception in Java – local variable. It can be annotated, but there is no way to get to local variables in the Reflection API, so the annotations on local variables are accessible merely before compilation.

IV. Conclusion

In previous sections we have introduced 6 different possible research directions in the field of attribute-oriented programming. First of the challenges was the research in the field of the alternatives to @OP metadata technique. The knowledge of viable alternatives for expressing and storing metadata about program elements and their comparison to annotations is useful in many situations, like in a case when a framework author needs to choose a metadata format for his/her framework. Annotations' restrictions are also an interesting research direction, but there are already methods and tools that address this issue. The challenge of finding a common API for annotations in different phases of program life is important for tools that need to process annotations in different phases. Absence of the common API requires writing multiple interfaces to the different APIs. The problem of annotating unnamed program elements is interesting, but motivation for solving it seems lower than the motivation for the rest of the challenges mentioned. Moving to other programming paradigms is less interesting too, since currently the mainstream paradigm is object-oriented programming anyway with challenges that need solution there. The question of when and how to use annotations is one of the hot issues. Methodologies and patterns of how to use annotations to deal with current programming and software design problems would be very useful for programmers.

In our future research we want to focus to the question of when and how to use annotations. We believe that this field is must current and still with too little advance. A related question to this is the problem of metadata alternatives to annotations. The question of alternatives is supposed to answer when to use alternatives instead of annotations. To answer these question we want to start by analysing alternatives and comparing them with the annotations. That should be able to help us with answering when to use annotations too. Then the future work is supposed to analyse more deeply the problem of annotations' usage by using annotations to preserve design decisions and store semantic informations about the code directly in the code. By focusing on the semantic and design annotations we narrow the field to managable problem.

Acknowledgment

This work was supported by VEGA Grant No. 1/0305/11 Co-evolution of the Artifacts Written in Domain-specific Languages Driven by Language Evolution.

References

[1] Allain A., 'Lambda Functions in C++11 - the Definitive Guide', CProgramming.com, http://www.cprogramming.com/c++11/c++11-lambda-closures.html, (2012)
[2] Cepa V., 'Attribute enabled software development: illustrated with mobile software applications', VDM Verlag, Saarbrücken, Germany, (2007)
[3] Cepa V., Mezini M., 'Declaring and Enforcing Dependencies Between .NET Custom Attributes', in Gabor Karsai & Eelco Visser, ed., 'GPCE' , Springer, pp. 283–297, (2004)
[4] ELCA, 'Xdoclet for more effective EJB development', POSMotivation for XDoclet, ZH, (2004)
[5] Greenfield J., Short K., 'Software Factories: Assembling Applications with Patterns, Models, Frameworks, and Tools' Wiley, Indianapolis, (2004)
[6] Java Community Process, 'JSR-000220 Enterprise JavaBeans 3.0', http://jcp.org/aboutJava/communityprocess/final/jsr220/, (2012)
[7] Java Community Process, 'JSR-000222 Java(TM) Architecture for XML Binding (JAXB)', http://jcp.org/aboutJava/communityprocess/mrel/jsr222/index.html, (2012)

[8] Java Community Process, 'JSR-000224 Java API for XML-Based Web Services 2.0', http://jcp.org/aboutJava/communityprocess/final/jsr224/index.html, (2012)
[9] JBoss Community, 'Hibernate', http://www.hibernate.org/, (2012)
[10] Mernik M., Heering J., Sloane A. M., 'When and how to develop domain-specific languages', ACM Computing Surveys (CSUR) 37 (4), pp. 316–344, (2005)
[11] Noguera C., Pawlak R., 'AVal: an extensible attribute-oriented programming validator for Java', Journal of Software Maintenance 19 (4), pp. 253–275, (2007)
[12] Nosáľ M., 'Overview of Literate and Elucidative Programming', SCYR 2012: Proceedings from conference: 12th Scientific Conference of Young Researchers, May 15th, 2012, Herľany, Slovakia, pp. 204–207, (2012)
[13] Pawlak R., 'Spoon: Compile-time Annotation Processing for Middleware', IEEE Distributed Systems Online 7, 11, (2006)
[14] Porubän J., Nosáľ M., 'Common Abstraction of Configuration from Multiple Sources', Acta Electrotechnica et Informatica, 11, 4, pp. 25–30, (2011)
[15] Ruska Š., Porubän J., 'Defining Annotation Constraints in Attribute Oriented Programming', Acta Electrotechnica et Informatica, 10, 4, pp. 89–93, (2010)
[16] SpringSource, 'Spring', http://www.springsource.org/, (2012)
[17] Tilevich E., Song M., 'Reusable enterprise metadata with pattern-based structural expressions', in Jean-Marc Jézéquel & Mario Südholt, ed., 'AOSD', ACM, pp. 25–36, (2010)
[18] XDoclet Group, 'XDoclet: Attribute-Oriented Programming for Java', http://xdoclet.sourceforge.net/, (2012)
[19] Wada H., Suzuki J., 'Modeling Turnpike Frontend System: A Model-Driven Development Framework Leveraging UML Metamodeling and Attribute-Oriented Programming', in Lionel C. Briand & Clay Williams, ed., 'MoDELS' , Springer, pp. 584–600, (2005)
[20] willCode4Beer, 'Annotations, the Good the Bad and the Ugly', willCode4Beer's blog, http://willcode4beer.blogspot.sk/2007/12/annotations-good-bad-and-ugly.html, (2007)

Case-Based Reasoning for Army Compositions in Real-Time Strategy Games

[1]*Martin ČERTICKÝ,* [2]*Michal ČERTICKÝ (4th year)*
Supervisor: [3]*Peter SINČÁK*

[1,3]Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic
[2]Dept. of Applied Informatics, FMPH Comenius University of Bratislava, Slovak Republic

[1]martin.certicky@tuke.sk, [2]certicky@fmph.uniba.sk, [3]peter.sincak@tuke.sk

***Abstract*—Over the years, there have been numerous successful applications of artificial intelligence techniques in the field of computer gaming. However, traditional graph-search based techniques often fail to perform at human level in real-time games with non-discrete game states. This fact encouraged the research of Case-Based Reasoning (CBR) and its applications to various aspects of computer game AI, especially in case of real-time strategies. We show how CBR can be used in the process of selecting the most effective army composition in a strategy game StarCraft, based on the game-related knowledge base designed by human experts (gamers).**

***Keywords*—Case-Based Reasoning, Agent, Real-Time Strategy, StarCraft**

I. Introduction

Players of adversarial computer games often need to adapt and react promptly and effectively to their opponent's unpredicted strategy. Most known artificial intelligence techniques have already been applied to computer games [11]. In this paper we will describe theoretical basics of an AI technique called Case-Based Reasoning (CBR) and show how it can be used to address one of the challenges from the field of game AI. CBR solves current, previously unseen, problems based on the solutions of similar past problems. The method includes the comparison process of current situation to previous similar cases. It chooses the most similar case from the past, and uses its corresponding remembered solution.

Different game genres present different challenges for artificial players (agents/bots). This work focuses on so-called Real-Time Strategy games (RTS). In RTS, as in other wargames, the participants position and maneuver units and structures under their control to secure areas of the map and/or destroy their opponent's assets. RTS games have shown to have huge decision spaces that cannot be dealt with search based AI techniques [11].

We will describe one possible application of CBR to playing RTS games, specifically for dynamic selection of the effective army composition in response to opponent's strategy. This is being done based on observation of opponent's actions and consequently comparing this knowledge with our case database. We have implemented an agent using CBR in this decision-making process within a real-time strategy game StarCraft: Brood War[1].

[1]StarCraft and StarCraft: Brood War are trademarks of Blizzard Entertainment, Inc. in the U.S. and/or other Countries.

The main goals of the paper is to present the practical application of CBR to an agent playing StarCraft and to demonstrate that this technique can help create intelligent, human-like behaviour in RTS games in general.

After the extensive overview of related work in section II, section III describes in detail how the CBR is used to solve our specific problem of selecting an optimal army composition. In section IV, we introduce our agent and give some insight into its implementation.

II. Related Work

Over the last years, case-based reasoning has grown from a rather specific and isolated research area to a field of widespread interest [1]. The number if its applications in various areas, including the game AI and opponent modelling, is rapidly growing.

Focusing only on the domain of game AI research and CBR applications relevant to this field, we were able to identify a considerable amount of published work. It therefore makes sense to categorize the games based on some kind of taxonomy, such as the one introduced by Aha, Molineaux and Ponsen in [2]. The games and corresponding CBR research is divided based on both the traditional player's viewpoint and on the degree of attracted research interest into following 7 categories:

A. Classic board games:

Typical board games like chess or checkers present a discrete, deterministic environment with two agents (players) affecting it in episodic turns. Several researchers have addressed classic board games, beginning with Arthur Samuels rote learning approach for playing checkers [15]. De Jong and Schultzs GINA instead memorized a partial game tree for playing Othello [4]. Chess has also been a popular topic. For example, Kerner described a method for learning to evaluate abstract patterns [10]. More recently, Powell et al.'s CHEBR learned to play checkers given only a paucity of domain knowledge [12].

B. Adventure games:

In adventure games, CBR has been used mainly for automated content generation. Fairclough and Cunningham described OPIATE [7], which uses a case-based planner and

constraint satisfaction to provide moves for a story director agent so as to ensure that characters act according to a coherent plot. Also, Daz-Agudo et al. described a knowledge-intensive approach [5] that extracts constraints from a users interactively-provided specification, uses them to guide case retrieval and adaptation, and then creates a readable plot using natural language generation techniques.

C. Team sports:

Team sport games provide a challenging environment for the problems of real-time multi-agent coordination and planning, but do not involve complicating dimensions common to strategy games, such as economies, research, and warfare. Quite popular instance of such game is the RoboCup Soccer [14]. Wendler and Lenz described an approach for identifying where simulated agents should move [19], while Wendler et al. reported strategies for learning to pass [20]. Gabel and Veloso instead used a CBR approach to select members for a team [9].

D. Real-time individual games:

Real-time games with single agent, such as first-person shooters, leave little space for CBR application, but there have been a few applications anyway. For example, Fagan and Cunningham focused on a plan recognition task - they acquired cases (state-action planning sequences) for predicting the next action of a human player [6].

E. Real-time god/management games:

Single player management games require agents to mainly deal with planning and plan adaptation tasks in possibly non-deterministic environment. Fascianos MAYOR system [8] learns from planning failures in Sim City[2], a real-time city management game. MAYOR monitors planning expectations and employs a causal model to learn how to prevent failure repetitions, where the goal is to improve the ratio of successful plan executions.

F. Discrete/turn-based strategy:

Complex turn-based strategy games, like Freeciv (open-source Civilization clone), require players to solve a number of distinctive sub-tasks. An example of applying CBR to such sub-task is Ulam et al.'s approach to defending the cities from attackers [16].

G. Real-time strategy:

RTS games usually focus on military combat (versus one or more adversaries), although they also include decision dimensions concerning tasks such as exploration, economic development, or research advancement in a non-deterministic, partially observable environment. Aha et al. used a case-based system CaT [2] to select offensive and defensive actions, and Weber with Mateas [18] used CBR to select a build order in Wargus (open-source Warcraft 2 clone). Cadena and Garrido presented the combined approach using the Fuzzy sets and CBR to deal with strategic and tactical management in StarCraft [3].

[2]Sim City is a trademark of EA International Ltd.

III. Our Work

Generally, in RTS games, a player frequently faces a number of "typical situations". Experienced players know how to respond to these situations in optimal way thanks to knowledge acquired by playing the game a lot.

Our artificial agent needed to have some kind of knowledge representation structure, that would be able to hold such information, and allow an agent to use it during the gameplay.

An intuitive choice of method for representing and reasoning about "typical situations" is CBR, where every such situation is considered an individual case.

The classic definition of CBR was coined by Riesbeck and Schank [13]:

"A case-based reasoner solves problems by using or adapting solutions to old problems."

Conceptually CBR is commonly described by the CBR-cycle (Fig. 1). This cycle comprises four activities (the four REs):

1) Retrieve similar cases to the problem description.
2) Reuse a solution suggested by a similar case.
3) Revise or adapt that solution to better fit the new problem if necessary.
4) Retain the new solution once it has beed confirmed or validated [17].

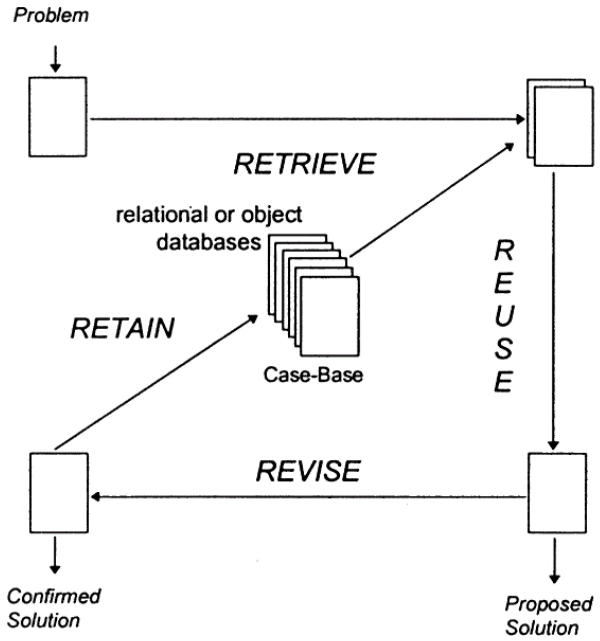


Fig. 1. Four activities of the CBR cycle as described in [17].

In our case, these problems (cases) are certain game situations. More specifically, a game situation is represented by a single *composition* of opponent's *army*.

Solutions to these problems are adequate counter-strategies, crafted by human players over time. A counter-strategy consists of our *desired army composition* and a collection of *desired upgrades* and *technologies*.

It is rare in RTS games that certain army composition has more than one equally good *"counter-composition"* (counter-strategy).

Generating such counter-compositions by some kind of rule-based system would be too difficult (near unrealizable in real

time), because of the complexity of StarCraft. Within every composition, there exists a large number of synergic effects and the overall effectivity against other compositions depends on too many different factors.

CBR allows us to abstract from the low-level game mechanics and in-depth origins of the effectivity of individual compositions, and take advantage of long-term experience of human players.

We are able to make use of predisposed database of game situations which we created. So far, we have not experimented with automatic modification of this database by an agent.

IV. Implementation

To test this approach, we implemented an agent playing a real-time strategy game StarCraft (Fig. 3). The agent was programmed in Java, using a JNIWBAPI interface, that allowed us to access in-game information and issue order to our units in real time, during the course of the game.

Simply put, our agent was constantly producing certain types of units during the game. The choice of unit types to produce is a result of evaluating the current game situation and selecting the adequate (most similar) case.

Each *case* is a vector consisting of doubles $\langle ratio, unitType \rangle$, where $unitType$ is a specific kind of opponent's unit and $ratio$ enumerates its percentage in his army.

A *solution*, corresponding to a case, is a similar structure. It also contains a collection of $\langle ratio, unitType \rangle$, but $unitType$s here correspond to units that we want to have in *our own army*. Additionally, a solution can also contain several $upgrades$ and $technologies$ that we want to research.

The case-solution list (database) is stored in simple text file, where every line contains one case and a corresponding solution.

Few examples can be found in Fig. 2 (note that the original lines are divided, and some unit types are omitted to fit the paper format).

```
15;Medic,20;Firebat,65;Marine---
          40;Zealot,40;Dragoon,20;HighTemplar
          @LegEnhancements,PsionicStorm

60;SiegeTank,40;Vulture---
          50;Zealot,40;Dragoon,10;Arbiter
          @LegEnhancements,StasisField

20;Ultralisk,15;Defiler,65;Zergling---
          30;Archon,20;DarkArchon,50;Zealot
          @MindControl,LegEnhancements
```

Fig. 2. Example cases with corresponding solutions. Solutions are divided by `"---"` string.

Our agent uses a simple *similarity function* to determine which case resembles the current game situation the most. Specifically, similarity function $S(A, B)$ compares two cases (army compositions) A, B and returns the percentage of units that they have in common. Let n be a total number of unit types in the game and let $ratio(A, i)$ denote the percentage of $i-th$ unit type in the case A. Function $S(A, B)$ is then defined as:

$$S(A,B) = \sum_{i=0}^{n} MIN\big(ratio(A,i), ratio(B,i)\big)$$

Fig. 3. In-game screenshot of our agent playing a StarCraft 1 vs. 1 match.

After we find the case that resembles current game situation the most, we use its solution as a parameter for our production function, which tells the agent what unit types (and in what ratio) to produce.

We evaluate game situations simply by observing the opponent's army composition. It is of course changing constantly during the game. Hence, in order to be successful, it is essential to properly monitor the game map (to "scout").

Consider the following example of army composition adaptation based on CBR: Thanks to scouting, we have noticed that the opponent started producing many air units (switched from *"100%zerglings"* to *"50%zerglings, 50%mutalisks"* army). Our own army (*"100%zealots"*) therefore becomes uneffective, since it cannot attack air units. However, using our similarity function, we determine that this situation resembles one of the typical cases from our database - the one with associated solution telling us to produce army effective against both air and ground enemy units (*"30%corsairs, 30%dragoons* and *40%zealots"*). When we tell our production function to produce this new composition, we should be able to deal with new opponent's army.

V. Conclusion & Future Work

Case-Based Reasoning offers a fine way to increase efficiency of intelligent agents playing strategy games. In future we are planning to create a system where our case database will be able to adapt by itself. This will be done based on game results and even on results of particular game sections (fights, economic situations etc.). We are also planning to include our CBR module in a student project *MontyBot*. The main goal of this project, solved by several students from different universities, is to create an intelligent StarCraft-playing agent using a wide variety of methods of artificial intelligence.

In our paper we described step-by-step the CBR technique, along with one of its possible application. We are certain this method has a great potential not only in this field.

References

[1] A. Aamodt, E. Plaza, *"Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches,"* AI Communications 7, 1994, pp. 39-59.

[2] D. W. Aha, M. Molineaux, M. Ponsen, *"Learning to Win: Case-Based Plan Selection in a Real-Time Strategy Game,"* Case-Based Reasoning Research and Development, Lecture Notes in Computer Science Volume 3620, 2005, pp. 5-20.

[3] P. Cadena, L. Garrido, *"Fuzzy Case-Based Reasoning for Managing Strategic and Tactical Reasoning in StarCraft,"* Advances in Artificial Intelligence, Lecture Notes in Computer Science Volume 7094, 2011, pp. 113-124.

[4] K. De Jong, A. C. Schultz, *"Using experience-based learning in game playing,"* Proceedings of the Fifth International Conference on Machine Learning, 1988, pp. 284-290.

[5] B. Daz-Agudo,P. Gervs, F. Peinado, *"A case based reasoning approach to story plot generation,"* Proceedings of the Seventh European Conference on Case-Based Reasoning, 2004, pp. 142-156.

[6] M. Fagan, P. Cunningham, *"Case-based plan recognition in computer games,"* Proceedings of the Fifth International Conference on Case-Based Reasoning, 2003, pp. 161-170.

[7] C. R. Fairclough, P. Cunningham, *"AI structuralist storytelling in computer games,"* Proceedings of the International Conference on Computer Games: Artificial Intelligence, Design and Education, 2004.

[8] M. J. Fasciano, *"Everyday-world plan use,"* Technical Report TR-96-07, The University of Chicago, Computer Science Department, 1996.

[9] T. Gabel, M. Veloso, *"Selecting heterogeneous team players by case-based reasoning: A case study in robotic soccer simulation,"* Technical Report CMU-CS-01-165, Pittsburgh, PA: Carnegie Mellon University, School of Computer Science, 2001.

[10] Y. Kerner, *"Learning strategies for explanation patterns: Basic game patterns with applications to chess,"* Proceedings of the First International Conference on Case-Based Reasoning, 1995, pp. 491-500.

[11] S. Ontañón, K. Mishra, N. Sugandh, A. Ram, *"Case-Based Planning and Execution for Real-Time Strategy Games,"* Lectuire Notes in Computer Science Volume 4626, 2007, pp. 164-176.

[12] J. H. Powell, B. M. Hauff, J. D. Hastings, *"Utilizing case-based reasoning and automatic case elicitation to develop a self-taught knowledgeable agent,"* In D. Fu and J. Orkin (Eds.) Challenges in Game Artificial Intelligence: Papers from the AAAI Workshop, 2004.

[13] C.K. Riesbeck, R. Schank, *"Inside Case-based Reasoning,"* Erlbaum, Northvale, NJ, 1989.

[14] J. Ruiz-del-Solar, E. Chown, P. G. Ploeger, *"RoboCup 2010: Robot Soccer World Cup XIV,"* Lecture Notes in Computer Science, Vol. 6556, 2011.

[15] A. Samuel, *"Some studies in machine learning using the game of checkers,"* IBM Journal of Research and Development, 3(3), 1959, pp. 210-229.

[16] P. Ulam, A. Goel, J. Jones, *"Reflection in action: Model-based self-adaptation in game playing agents,"* In D. Fu and J. Orkin (Eds.) Challenges in Game Artificial Intelligence: Papers from the AAAI Workshop, 2004.

[17] I. Watson, *"Case-based reasoning is a methodology not a technology,"* AI-CBR, University of Salford, Salford M5 4WT, UK, 1999.

[18] B. G. Weber, M. Mateas, *"Case-Based Reasoning for Build Order in Real-Time Strategy Games,"* In Proceedings of the Fifth Artificial Intelligence for Interactive Digital Entertainment Conference (AIIDE), 2009, pp. 106-111.

[19] J. Wendler, M. Lenz, *"CBR for dynamic situation assessment in an agent-oriented setting,"* In D.W. Aha and J.J. Daniels (Eds.), Case-Based Reasoning Integrations: Papers from the AAAI Workshop, 1998.

[20] J. Wendler, G. A. Kaminka, M. Veloso, M. *"Automatically improving team cooperation by applying coordination models,"* In B. Bell and E. Santos (Eds.) Intent Inference for Collaborative Tasks: Papers from the AAAI Fall Symposium, 2001.

Components of the Web Interface for the SLAmeter Tool

[1]*Maroš VASIĽ*, [2]*Adrián PEKÁR (2nd year)*
Supervisor: [3]*Liberios VOKOROKOS*

[1,2,3]Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

[1]maros.vasil@student.tuke.sk, [2]adrian.pekar@tuke.sk

***Abstract*—This paper deals with the SLAmeter tool, which provides an alternative solution to the existing network traffic monitoring tools. By its design, SLAmeter offers a reliable option to evaluate the Service Level Agreement (SLA) for tracking the fulfillment of the specified conditions between the subscriber and the provider. Particular attention is given to the tool's web interface and its structure, as well as to the description of the SLAmeter's main components. This paper also discusses the technology framework standing behind the particular components and the entire application structure.**

***Keywords*—SLA, IPFIX, Modular Web Application, Traffic Monitoring**

I. Introduction

In the recent few years the technologies of computer networks went through a significant breakthrough. The rapid development of the Internet has contributed to the informatization of the whole society. Consequently, a large volume of various information mix has to be transferred between the end points of the communication. This information mix — which usually consist of video, voice and data — is also known as network traffic. While the Internet and its services are used during every day life in an increasing routine, the management, utilization and security [12], [13] of computer networks with such a huge volume of network traffic is becoming an ever demanding task. According to [9], network traffic monitoring and properties measurement represents an essential tool for keeping the functionality of today's networks on an acceptable level.

In the context of Internet traffic disruptions, many different analyzers had been created that can indicate various issues of the network. The advantage of these analyzers is, that they can help to fix or limit the errors in the network's management and improve its overall functionality. One tool that collects and analyzes the traffic parameters of the network is the SLAmeter tool. SLAmeter is being developed by the MONICA research group in the Computer Networks Laboratory of the Technical University of Košice. By processing the data on the input, SLAmeter provides various information about the quality of Internet connection in a form of charts, tables or listings. SLAmeter provides an alternative solution to the widespread 'Speedmetres' [7] too.

The goal of this paper is to introduce the design and implementation details of some components of the SLAmeter's web interface, which is located at the highest level of the tool's architecture. The components are designed to work independently. By its modular design, extra functionalities and other components can be easily added or removed.

In the following sections the design of the SLAmeter's web interface will be presented, starting with a brief description of the SLAmeter itself, continuing with an overview of the web interface and its main components, up to the depiction of the development of these components. The last section of this paper draws a conclusion and some future directions.

II. Basic Concept

The main objective was to create a straightaway, easy to use web interface for the SLAmeter tool. The basic idea of the web interface (for displaying the resulting HTML code) was a framework, which is independent from its components. The configuration of the components are possible only from a configuration file, which defines the used one components.

The framework itself is simple, load avoided and does not contain any code related to the components. The web interface provides the following forms of output:

- Chart – the received data are displayed by various graphical charts (Fig. 1). Most of the data will be measured and processed according to a time scale, so one of the chart's axes is usually a timestamp and the other a measured value.
- Table – another method for displaying the output is by the means of tables. As shown in Table I, some statistical data will be displayed by tables.
- Text – for providing a meaningful result, some data do not need any complicated approach for displaying the output. So some data will be shown just as simple text. The example in Fig. 2 is depicting a specific list of SLA class values.

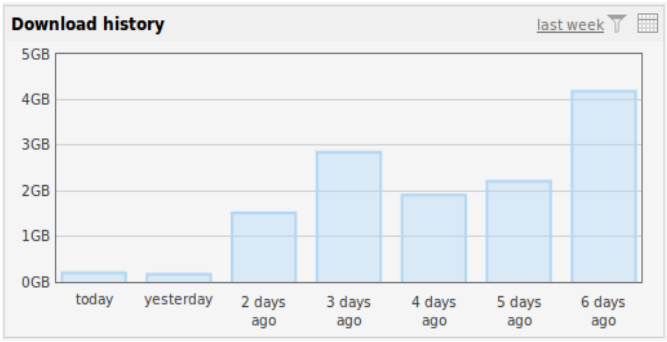


Fig. 1. Example of a chart displayed by the SLAmeter tool.

TABLE I
EXAMPLE OF A TABULAR OUTPUT FORM OF THE SLAMETER TOOL.

Remote IP Address	Local Port	Remote Port	Protocol	Downloaded	Uploaded
123.123.123.123	4578	80(HTTP)	TCP	10.4 MB	654 kB
192.168.1.4	5421	80(HTTP)	TCP	15.8 MB	267 kB
222.222.333.232	51234	25(SMTP)	TCP	5.7 MB	632 kB

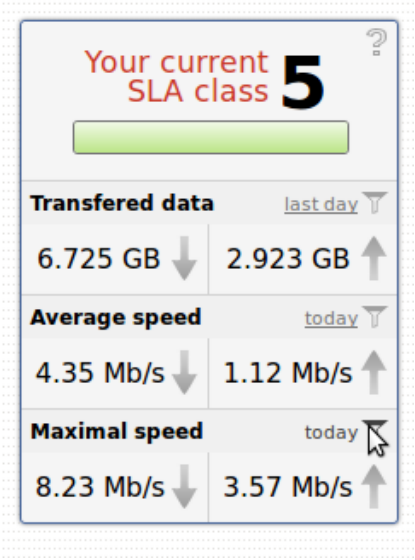


Fig. 2. The example of the SLAmeter's text output format.

III. THE DESIGN OF THE SLAMETER TOOL

The architecture of the SLAmeter tool is divided into several parts. Each of these parts performs some sub-tasks that are necessary for the errorless operation and functionality of the SLAmeter tool. These parts are as follows:

- Exporter – captures network traffic and classifies it into flows [10], [14]. Based on the captured data it creates flow records and using the IPFIX protocol [2] exports them to the Collector [2].
- Collector – collects flow records obtained from the exporters. These records are subsequently stored in a database where the data can be accessed for further processing by the Analyzer [8].
- Analyzer – represents the user front-end of the SLAmeter tool. Its main aim is to analyze and evaluate the information from the flow records collected by the Collector. It calculates the measured parameters and displays them for the user [1], [11]. The Analyzer is further divided into two parts:
  - Evaluator – based on the requirements, it processes IPFIX records and creates statistical and analytical data. Its architecture is modular, so it is composed from a framework and the components that are embedded into it. The components calculate individual data and provide them in a simple numerical form. The responses for requests are formed by the Evaluator into JSON objects [1].
  - Web interface – the web interface is highly modular which means, that it consist of a framework and some components that are embedded into it. HTML code of the resulting page consists from fragments of HTML code for each component. The framework subsequently composes these data into the resulting format [11]. The entire architecture is shown in Fig. 3.

IV. THE IMPLEMENTATION OF THE WEB INTERFACE

A. Used Technologies

To develop the components of SLAmeter web interface, the Python programming language [6] with the Django framework [4], [5] was used. For drawing charts, the JavaScript programming language with the Flot library was selected [3].

B. Implementation of the Web Components

Each component consists of two separate files. The first file is written in the Python programming language. It contains two methods. The first method is a `GetRequest` method, which contains the requirement for the particular component with the requested contents. The second is a `getBlockHTML` method, that processes the JSON response and transfers the necessary data into the template file. The example of these two methods is provided below.

```
class MaximalSpeed():

@classmethod
def getRequest(cls):
  # receive data from filter and
  create json for evaluator
  currentmillis = int(round(time.time()
  *1000))
  yesterdaymillis = currentmillis
  -86400000

  return "{\"name\":
  \"MAXIMUMDOWNLOADUPLOAD\",
        \"request\":
      {\"networkpoint\": 1,
        \"timefrom\":
        "+str(yesterdaymillis)+",
           \"timeto\":
           "+str(currentmillis)+",
           \"ipAddress\":
           \"147.232.55.129\" }}"

@classmethod
def getBlockHTML(cls, jsonText):
  #receive data from json, create object
  and send html code back
  jsonRes = json.dumps(jsonText[
  "response"])
  jsonTitle = json.dumps(jsonText["name"])
  responseTitle = json.loads(jsonTitle)
```

```
  responseData = json.loads(jsonRes)

  t = loader.get_template('maximalspeed
  .html')
  c = Context({
      'title': responseTitle,
      'content: responseData,
      })
return t.render(c)
```

As mentioned above, a second file is a file with the template. In this file is located the command that extends the `block.html` file. In the case, that the component does not receive any response, this file provides the specific name of the component with a default value (content).The template contains another three sets of blocks. These are the block with the title of component, the block containing the values from the JSON response, and the block, that provides the modes of the component. The name of these blocks are "Basic", "Advanced", and "Simple". Simple mode displays the component, that is located on the sidebar. An example of the template file is provided below.

```
{% extends block.html %}
{% block title %}
{{ title }}
{% endblock %}
{% block class %}
basic
{% endblock %}
{% block content %}
<div class="tool-box"></div>
<div id="hd-plot" class="plot"></div>
<script type="text/javascript">
GRAPH.graphs.downloadHistory.plot(
"#hd-plot",[{{ content }}]);
</script>
{% endblock %}
```

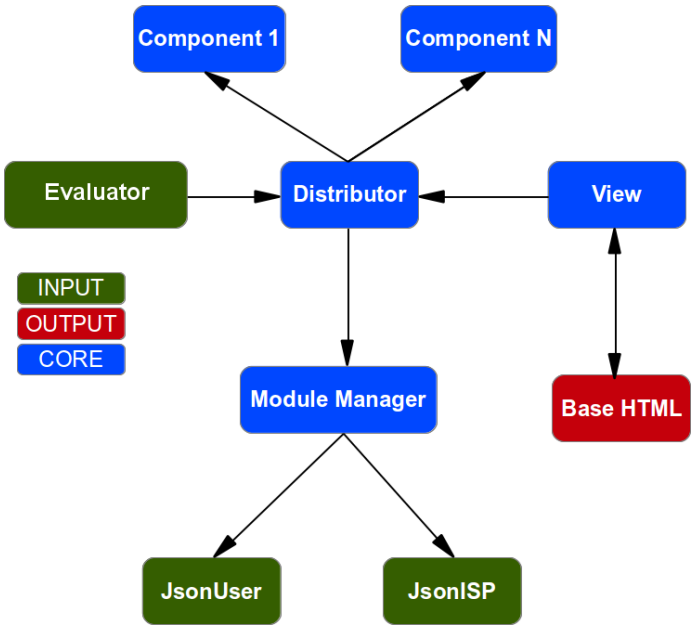


Fig. 3. The architecture of the SLAmeter web interface

V. Conclusion

This paper was devoted to the SLAmeter tool, which provides an alternative solution to the existing network traffic monitoring tools. By its design, SLAmeter offers a reliable option to evaluate the Service Level Agreement (SLA) for tracking the fulfillment of the specified conditions between the subscriber and the provider. The paper provided a description of the SLAmeter tool, as well as an overview of the web interface and its main components with the depiction of the development of these components.

The components were developed using the Python programming language with the Django framework using Pycharm environment. Django uses the MVC architecture, which serves for the separation of the application and presentation layer. The presentation layer was created in HTML language with Django markup.

Future work should be aimed at the development of new modules and components, which would extend the toolâĂŹs applicability in new domains, e.g. usage-based accounting, security analysis or real-time data evaluation.

VI. Acknowledgment

This work was supported by the Slovak Research and Development Agency under the contract No. APVV-0008-10 (50%). This paper is also the result of the project implementation: Competency Centre for Knowledge Technologies Applied in Innovation of Production Systems in Industry and Services, ITMS: 26220220155, supported by the Research & Development Operational Programme funded by the ERDF (50%).

References

[1] M. Antl, "Framework for the evaluator and web interface of the slameter tool," Master's thesis, Technická univerzita v Košiciach, Letná 9, 042 00, Košice, 2012.

[2] B. Claise, "Specification of the IP Flow Information Export (IPFIX) Protocol for the Exchange of IP Traffic Flow Information," RFC 5101 (Proposed Standard), Internet Engineering Task Force, 2008.

[3] Flot, "Attractive javascript plotting for jquery," 2013. [Online]. Available: http://www.flotcharts.org/

[4] A. Holovaty and J. Kaplan-Moss, *The Definitive Guide to Django: Web Development Done Right*, ser. Apresspod Series. Apress, 2009.

[5] A. Hourieh, *Learning Website Development with Django*. Packt Publishing Ltd., 2008.

[6] M. Lutz, *Learning Python: Powerful Object-Oriented Programming*, ser. Learning Python. O'Reilly Media, 2009.

[7] Ookla, "Speedtest.net," 2013. [Online]. Available: http://www.speedtest.net/

[8] A. Pekár, "Optimization of the collecting process for the basicmeter tool," Master's thesis, Technická univerzita v Košiciach, Letná 9, 042 00, Košice, 2011.

[9] A. Pekár, "Modelovanie a návrh systémov pre monitorovanie siet'ovej prevádzky," 2013.

[10] J. Quittek, T. Zseby, B. Claise, and S. Zander, "Requirements for IP Flow Information Export (IPFIX)," RFC 3917 (Informational), Internet Engineering Task Force, 2004.

[11] M. Vasil', "Components of web interface for the slameter tool," Master's thesis, Technická univerzita v Košiciach, Letná 9, 042 00, Košice, 2012.

[12] L. Vokorokos, N. Ádám, and A. Baláž, "Application of intrusion detection systems in distributed computer systems and dynamic networks," in *In Proceedings of Computer Science and Technology Research Survey, (CST'08)*, 2008, pp. 19–24.

[13] L. Vokorokos, A. Kleinová, and O. Látka, "Network security on the intrusion detection system level," in *In Proceedings of the 10th IEEE International Conference on Intelligent Engineering Systems, (INES'06)*, 2006, pp. 270–275.

[14] T. Zseby, E. Boschi, N. Brownlee, and B. Claise, "IP Flow Information Export (IPFIX) Applicability," RFC 5472 (Informational), Internet Engineering Task Force, 2009.

Concept definition for process discovery purposes

[1]*Cecília HAVRILOVÁ (1st year)*
Supervisor: [2]Ján PARALIČ

[1, 2] Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

[1]cecilia.havrilova@tuke.sk, [2]jan.paralic@tuke.sk

***Abstract*— The aim of this paper is to provide a theoretical description of the formal algorithm concept, designed for the purpose of process mining - one of the tasks of knowledge discovery in databases (KDD). CRISP-DM methodology was chosen as a standard of KDD process. Within each phase of the methodology, particular steps of knowledge discovery are analogically specified for the purpose of process mining, i.e. definition of the target task, event logs collection, event logs preparation and transformation to a suitable form, choosing and applying the selected algorithm, evaluation of results, in terms of established process mining task and implementation of obtained knowledge in practice.**

***Keywords*— CRISP-DM, event logs, KDD, process mining**

I. INTRODUCTION

Data and information are now the most valuable assets in many companies. Organizations around the world are increasingly aware of the strategic value of information, stored in their information systems, whether for business information systems, management information systems, business intelligence or others.

The ability to transform data obtained from everyday transactions into valuable information, leads to significant improvements in management and decision-making. For proper tactical and strategic decision making, it is not necessary to be in possession of large data sources, but it is important to know, how to analyze the available data and derive new knowledge with significant impact [1].

There are several possibilities of using data mining techniques and analyses of data sets. One of these possibilities is the process mining approach.

The aim of process mining is to discover, monitor and improve real processes by extracting knowledge from event logs that occur in information systems [2].

There are several areas of process mining techniques utilization. One of these areas is for example monitoring and mining in healthcare processes [3]. Another significant process mining application is fraud mitigation, described in detail in the paper by Jans, M. et al. [4]. Another process mining approach with the aim on redesigning the embedded business process of an enterprise application was presented by van Beest and Maruster [5].

Our aim is to redesign a formal concept of knowledge discovery algorithm into the context of process mining and describe particular steps of this algorithm. This approach can be used for different KDD tasks in context of process mining.

II. KNOWLEDGE DISCOVERY

Process mining is closely related to data mining. Whereas classical data mining techniques are mostly data-centric [6], process mining is process-centric.

Knowledge discovery in databases is iterative and interactive process of semi-automatic extraction of knowledge from databases [7]. This process consists of several steps. There are a lot of different definitions of KDD process, describing particular steps of this process at multiple levels of detail. Some efforts to unify these definitions can be found in general methodologies of KDD process. The most famous are SEMMA, 5A and CRISP-DM methodologies. The last mentioned - CRISP-DM, belongs to most commonly used. This methodology was conceived in 1996 as an industry tool and application neutral standard methodology for defining and validating data mining process [8]. CRISP-DM defines KDD process by following six phases (Fig. 1.).

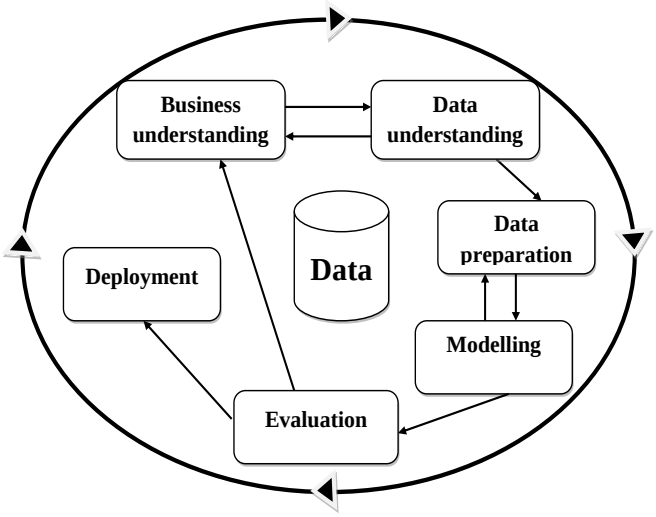


Fig. 1. CRISP-DM methodology

Description of CRISP-DM phases:

1. Business (problem) understanding - it is an initial phase, which focuses on definition of data mining (DM) task from a business perspective.
2. Data understanding - includes data collection, data description and verification of data quality.

3. Data preparation - preparation of data into suitable form.
4. Modeling - this phase includes choice of modeling techniques and their application.
5. Evaluation - model evaluation and its compliance with the results of phase 1.
6. Deployment - application of created model in practice

The main advantage of this methodology lies in its universal applicability on any data mining tasks, i.e. classification or prediction tasks, clustering tasks, association rules discovery, etc.

III. ALGORITHM PROPOSAL

Proposed algorithm is based on CRISP-DM methodology. It is a universal formal algorithm of KDD adapted to the context of process mining tasks. Knowledge obtained as a result of mining the event logs illustrates the ongoing process in form of process model, which can be represented for example as a Petri net. The proposed algorithm is depicted in Fig. 2. Particular steps of this algorithm are described in the following chapters.

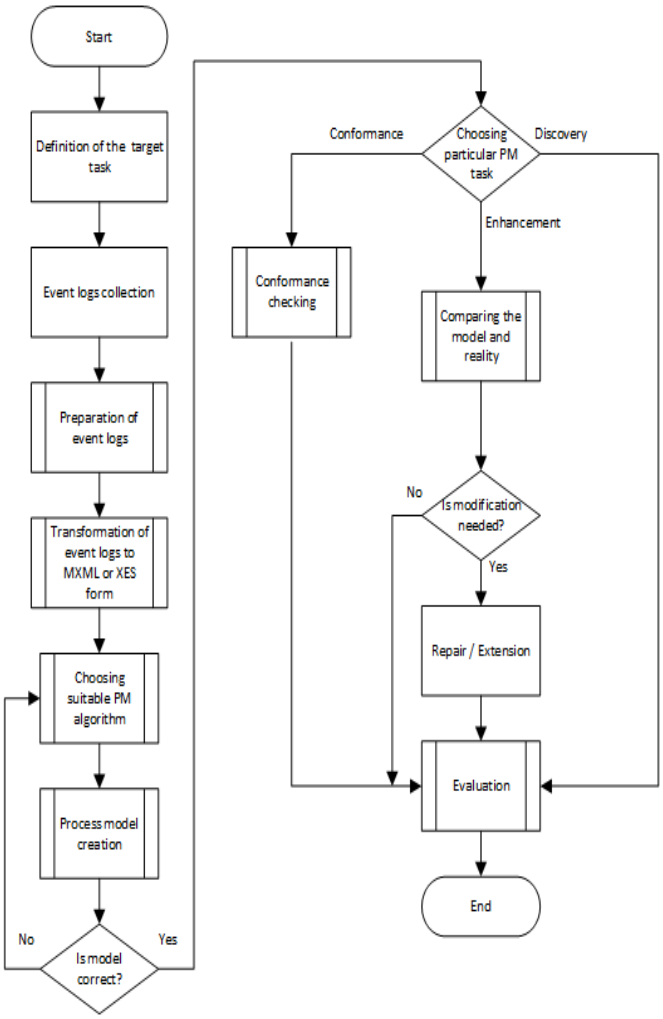


Fig. 2. Proposed formal algorithm of KDD in context of process mining tasks

A. Definition of target task

Software processes, ongoing in enterprise information systems, are daily recorded and kept. This enables to monitor and subsequently coordinate and simplify all these processes [9].

The final task definition may differ in dependence with the purpose of results usage expectations. Final task can be defined as, for example the optimization of the manufacturing process with aim on minimizing of losses volume, time consumption, investments volume and workforce volume or maximizing the profit. Other possible application of process mining is the possibility of atypical patterns detection, which do not meet the criteria of a previously defined normal workflow. Also a possible case of process mining usage is the analysis of relations between the actors of a process, i.e. definition of a social network in frame of an organization. Process mining techniques can also be used for weak and strong spot detection in already executed processes, thus creating a set of recommendations.

B. Event logs collection

Process mining is generally based on analysis of data generated by already executed activities mediated by information system(s) generating events, denoted as event logs. An event is used to characterize the dynamic behavior of a process in terms of identifiable, instantaneous actions, or decision making upon the next activity that is to be performed [9].

Each event in such a log refers to an *activity* i.e., a well-defined step in some process [10]. Event log can contain different attributes, e.g. activity ID, time stamp (time of the event execution), time length of the event, description of the executed activity, activity owner, etc. All these attributes play significant role by means of meeting the conditions for task completion, defined in subchapter A. Event logs' collecting is usually carried out in the background, by company's enterprise resource planning system (ERP), other information systems or applications, created only for the particular purpose of event log collection. It is possible to periodically archive event logs of manufacturing lines, software processes, web services, clouds and many others.

C. Preparation of event logs

Data preparation is very difficult and time-consuming process. Within KDD, data preparation process consists of the following phases [7]:

1. data cleansing
2. data integration from various sources
3. data transformation into suitable form
4. data construction

In the context of process mining, the key role of data preparation is concentrated on data integration and cleansing. Data scope, volume and quality have significant impact on the process model definition. This is the main reason why it is essential to have all event logs available. Data cleansing aims at removing noisy event logs, i.e. removing incomplete or inconsistent event logs or event logs with poor information value.

D. Transformation of event logs to MXML or XES form

Event logs of ERP systems are usually archived as documents, saved in internal databases of organizations. The extraction of logs into suitable form requires data transformation into suitable format, e.g. MXML or XES form. These two forms are required as a standard in different

process mining software tools. There are multiple solutions how to extract event logs in MXML or XES form from various data sources, for example using tools such as ProMimport, Nitro, XESame etc.

E. *Choosing suitable process mining algorithm*

Phase of modeling includes choosing suitable algorithm for the purpose of event logs analysis and also for the process of model creation. The goal of this search is to find the process model that fits the data in the workflow log the best [11]. Since, there are a lot of algorithms used for process mining tasks [12]; it can be difficult to choose a suitable one. The manual selection usually depends besides the target discovery task also on properties of event logs, for example noise and parallel structure, incomplete workflow logs, duplicate tasks, invisible tasks, etc.

One of the general process mining algorithms is the α-algorithm. This algorithm discovers a process model in form of Petri net, from the local dependence between events [13]. Other algorithms are based on heuristic or genetic approaches [14], [15]. Another approach is based on neural nets or cluster analysis [16]. There is also a possibility to use techniques based on data mining approach and also Markovian approach.

F. *Process model creation*

The process model creation is based on application of chosen process mining algorithm on event logs. Created model can be visualized. For example α-algorithm produces a Petri net but it is easy to convert the resulting model into a BPMN model, BPEL model, or UML Activity Diagram [17].

G. *Is model correct?*

Validation and verification of a process model can also be represented by the phase of evaluation. Verification typically relies on an algorithmic analysis of the discovered process model, i.e. if it relates to the envisaged internal process model. Validation requires also the consultation of the internal specification and discussion with process stakeholders [18].

H. *Choosing particular process mining task*

Event logs can be used to conduct three types of process mining: discovery, conformance and enhancement [17].

a) Discovery - the first type of process mining task, process discovery, is used for automatic identification of process structure derived from the event logs of this process. Created process model can be represented afterwards for example as a Petri net (using α-algorithm).

b) Conformance - the second type of process mining task - *conformance checking* is used to check if reality, as recorded in the log, conforms to the real model and vice versa [17].

c) Enhancement - the third type of process mining task is enhancement. The aim of this approach is to *compare the model and reality* and then modify an existing process by using information from event logs analysis. If model obtained by this analysis does not correspond with an existing model, the existing *model can be modified by:*

a. reparation of this model

b. extension of this model

I. *Process model evaluation*

Evaluation phase is considered the concluding and also the most important part of the process and results understanding. It is realized using comparison between the source process model and the resulting process model. In case of process discovery tasks, the evaluation is represented by a comparison between the actually implemented business model or workflow that is already being used for the particular problem and the process model resulting from the process mining. We assume that in the future it may be also possible to use various metrics including metrics known form graph theory and network flows.

IV. Implementation Proposal

The implementation process is fairly straightforward. The process begins with data collection, goes through processing, using designated tools and after using process mining algorithms it reaches the implementation phase, which concludes process discovery in deployment and usage stage. Described implementation proposal is depicted in Fig. 3.

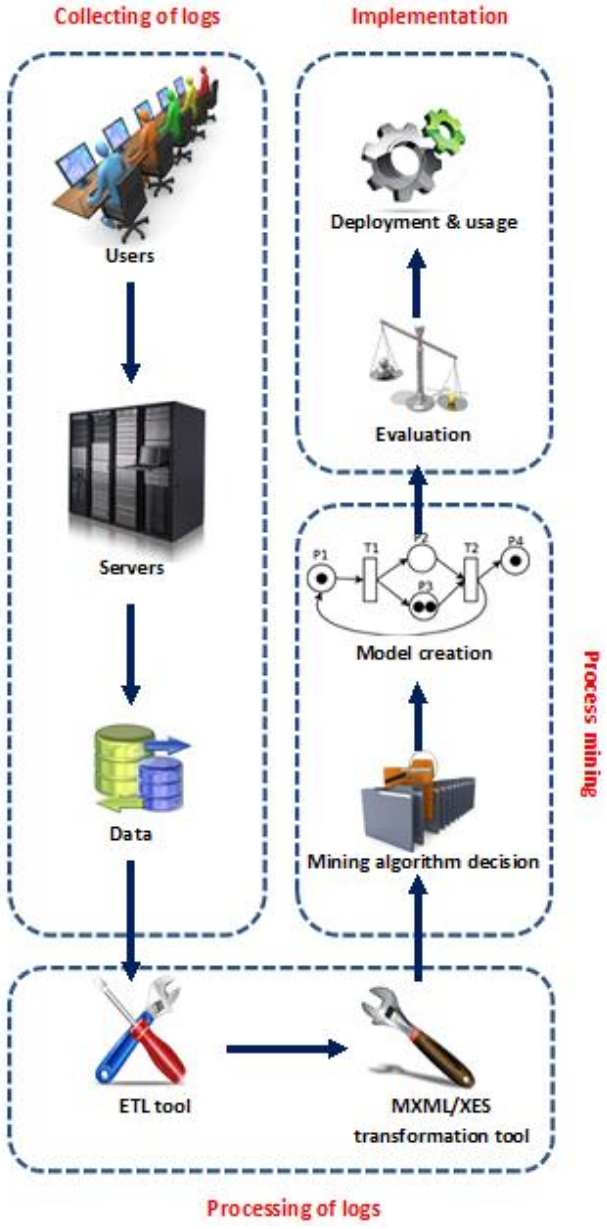


Fig. 3. Implementation proposal scheme

V. Conclusion

In this paper we have shown a general overview of an emerging field of research, the process mining. We have proposed a concept for carrying out process mining related tasks according to KDD process and used CRISP-DM methodology as a template. The main asset of this paper is the concept that can be used for further research of methods suitable for process mining activities and process mining related problems. Since the respective area is rather undiscovered, it is not yet possible to evaluate the usefulness of the proposed concept. This evaluation will be a task for further research in the future.

Acknowledgment

The work presented in this paper was partially supported by the Scientific Grant Agency of Ministry of Education, Science, Research and Sport of the Slovak Republic and the Slovak Academy of Sciences under grant No. 1/1147/12 (50%) and by the Slovak Cultural and Educational Grant Agency of the Ministry of Education, Science, Research and Sport of the Slovak Republic under grant No. 065TUKE-4/2011 (50%).

References

[1] HAVRILOVÁ, C., BABIČ, F.: Financial data analysis using suitable open-source Business Intelligence solutions. In: SAMI 2013: IEEE 11th International Symposium on Applied Machine Intelligence and Informatics: Herľany, Slovakia. - Budapest: IEEE, 2013 P. 257-262. - ISBN 978-1-4673-5927-6

[2] IEEE TASK FORCE ON PROCESS MINING, "Process Mining Manifesto," Proc. Business Process Mining Workshops, Lecture Notes in Business Information Processing,Springer, 2011.

[3] MANS, R.S. et al.: Process Mining in Healthcare: Data Challenges when Answering Frequently Posed Questions. In: Process Support and Knowledge Representation in Health Care - BPM 2012 Joint Workshop, ProHealth 2012/KR4HC 2012, Tallinn, Estonia.

[4] JANS, M. et al.: A business process mining application for internal transaction fraud mitigation. In: Expert Systems with Applications, Volume 38, Issue 10, 15 September 2011, Pages 13351–13359

[5] VAN BEEST, N.R.T.P, MARUSTER, L.: A Process Mining Approach to Redesign Business Processes - a case study in gas industry. In: SYNASC 2007: IEEE Ninth International Symposium on Symbolic and Numeric Algorithms for Scientific Computing: Timisoara, 2007 P.541-548. - ISBN 978-0-7695-3078-8

[6] HAND, D., MANNILA, H., AND SMYTH, P. 2001. *Principles of Data Mining*. MIT Press, 2001, Cambridge,MA., 546 pp. - ISBN 978-0-26-208290-7.

[7] PARALIČ, J.: Knowledge discovery in databases (Objavovanie znalostí v databázach) - Košice : Elfa, - 2003. - 80 pp. - ISBN 80-89066-60-7.

[8] KADAV, A., KAWALE, J., MITRA, P.: Data Mining Standards. In: http://www.datamininggrid.org/wdat/works/att/standard01.content.08439.pdf

[9] COOK, J.E., WOLF, A.: Discovering Models of Software Processes from Event-Based Data. In: ACM Transactions on Software Engineering and Methodology, Vol. 7, No. 3, July 1998, p. 215–249.

[10] W.M.P. VAN DER AALST: Process Mining: Overview and Opportunities. In: ACM Transactions on Management Information Systems, Vol. 3, No. 2, Article 7, Publication date: July 2012.

[11] W.M.P. van der Aalst, A.J.M.M. Weijters: Process mining: a research agenda. In: Computers in Industry 53, 2004. P. 231-244.

[12] TIWARI, A., TURNER, C.J., MAJEED, B.: A review of business process mining: state-of-the-art and future trends. In: Business Process Management Journal, Vol. 14 No. 1, 2008, pp. 5-22

[13] YUE, D. et al: A Review of Process Mining Algorithms. In: BMEI, 2011: IEEE Business Management and Electronic Information International Conference, Guangzhou. p. 181- 185. ISBN 978-1-61284-108-3

[14] LI, Ch., REICHERT, M., WOMBACHER, A.: Discovering reference models by mining process variants using a heuristic approach. Business Process Management, 2009, p. 344-362

[15] VAN DER AALST, W.M.P., DE MEDEIROS, A.K.A., WEIJTERS, A.J.M.M.: Genetic Process Mining. In: Ciardo, G., Darondeau, P. (eds.) ICATPN 2005. LNCS, vol. 3536, p. 48–69. Springer, Heidelberg (2005)

[16] SONG, M., GUNTHER, C.W., VAN DER AALST W.M.P.: Trace Clustering in Process Mining, BPM Workshops (2008)

[17] VAN DER AALST, WIL M. P.: Process Mining - Discovery, Conformance and Enhancement of Business Processes. 1st Edition, XVI, 352 p.. ISBN 978-3-642-19345-3

[18] MENDLING, J.: Empirical Studies in Process Model Verification. In: Transactions on Petri Nets and Other Models of Concurrency II: Special Issue on Concurrency in Process-Aware Information Systems. Berlin, Germany: Springer-Verlag, 2009, pp. 208–224.

Construction and Operating System of Robosoccer Agents

[1]*Michal KOPČÍK,* [2]*Radoslav BIELEK (1st year)*
Supervisor: [3]*Ján JADLOVSKÝ*

[1,2,3]Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

[1]michal.kopcik@student.tuke.sk, [2]radoslav.bielek@tuke.sk, [3]jan.jadlovsky@tuke.sk

***Abstract*—This paper presents the results of the continuous work on the robot constructed mainly for the robotic football purposes. Most of the article focuses on the robot's hardware as it is the most developed part so far, but we also bring the draw of the control software and signal processing with the formulas for the computation of motor's speed and torque and the robot's position. The communication and software parts are also briefly presented.**

***Keywords*—Mobile robot, robotic football, MiroSot, microcontroller, signal processing, position calculation**

I. Introduction

The main purpose and motivation to start the work on this robot was mainly to use it in the field of Robotic Football and to challenge other robotic football teams.

A. Robotic football

The field of robotic football includes several fields for example robotics, motion control and artificial intelligence with focus on planning, strategy and optimization. Whole field of "robosoccer" is organized by the FIRA – Federation of International Robot-soccer Association.

There are several leagues and modification of this scientific sport according to the robot's shape and weight, one category even without the robots, using only competing strategies in the simulated environment.

We decided to contribute in the modification called MiroSot (Micro Robot World Cup Soccer Tournament) in which all matches are played by two teams, each consisting of three robots (one of them can be goalkeeper, but can contribute also as a regular player). Robots have to be fully automated, but there are three people allowed to be on the stage: manager, coach and trainer. Computational power for the match is focused in one host computer mainly dedicated to the vision processing. Robot's dimensions are limited by 7.5 x 7.5 x 7.5 cm excluding robot's antenna. Game is played with and orange golf ball on the 400 x 280 cm or 220 x 180 cm large pitch according to the league modification (Large or Middle) [1].

Markedly achievements in MiroSot were obtained by the group called TUKE Robotics from Department of Production System and Robotics, Faculty of Mechanical Engineering, Technical University of Košice. They are multiple euro champions (2006 - 2009) and also world champions for the year of 2010 [2].

B. Other usage of robot

Even though the robot is constructed mainly for the purposes of the robotic football it can be used also in the MicroMouse competition which consists of solving the 16 x 16 maze. Events are held worldwide and demands usage of fully autonomous robots [3].

Last, but not least is the usage of the complete robot in the class, especially in the subjects of motion control and artificial intelligence.

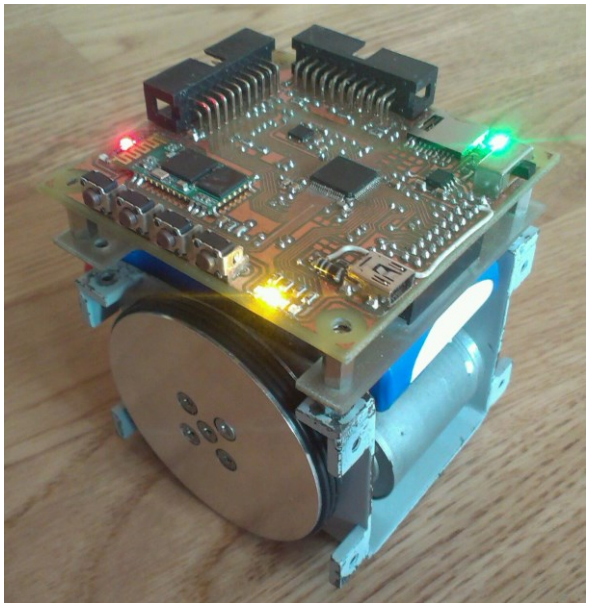


Fig. 1. Prototype of the robot.

II. Mechanics

Mechanics of the robot consists of chassis, wheels and motors. Since the robot is constructed with the use of differential drive, there are used two same motors.

A. Chassis

Chassis are made of 1.5mm steel plate. Wheel bearings are placed in the housings welded to the frame. For the material used, the frame is very durable, but also quite heavy.

B. *Wheels and gears*

Wheels are made of aluminum, they are 10 mm thick and the wheel diameter with tire is 62 mm. Tire of the wheel consists of three 2 mm O-rings placed in the groove on the wheel. Gear that connects wheel with the motor is located inside wheel and they are connected together using screws. Ratio of used gears is about k = 0.1.

C. *Motors*

Motors determine robot's performance such as maximum acceleration, maximum speed, current consumption, etc. Motors used on the robot are Mabuchi RS-380SH with nominal voltage of 7.2 V. Maximum speed and acceleration of the robot can be simply calculated from equation (1) and (2). Description of symbols used is presented in table 2.

$$v_{nom} = \frac{\omega_{nom} \times \pi \times d \times k}{60} \quad (1)$$

$$a_{nom} = \frac{2 \times \tau_{nom}}{d \times m \times k} \quad (2)$$

Acceleration and maximum speed depends on the mass of the robot (m = 0.55 kg) and gear ratio (k = 0.1). Motor driver can provide continuous current about 6 A to motors. At current 6 A motors provide torque about τ_{nom} = 0.04 N.m^{-1} per motor so derived from the equation (2) theoretical maximum acceleration is about a_{nom} = 26 m.s^{-2}. Maximum speed of the robot calculated from equation (1) for speed of the motor driven at 6 A is about v_{nom} = 3.75 m.s^{-1}. These values are only theoretical and doesn´t count externalities such as friction in gearbox and friction between tire and surface. Real speed of the robot is about 1/3 lower than calculated nominal speed and acceleration highly depends on material of the tire and surface. In real conditions maximum acceleration is about 5 m.s^{-2}.

III. Electronics

Whole electronics of the robot consists of control board, power board and two sensor boards. Boards are populated mostly by SMD components and PCB's (Printed Circuits Board) are homemade and hand soldered to reduce cost of the prototype.

A. *Control board*

Essential parts of the control board consist of 32-bit microcontroller and communication Bluetooth module. This board also contains MicoSD card socket, USB connector, gyroscope and accelerometer sensors, some buttons and LEDs for easier program debugging and 20-pin connector to connect extra sensors, actuators or digital camera.

Microcontroller used on this board is from a family of 32-bit Flash microcontrollers based on the ARM core Cortex-M3.

Control board is connected to the power board via 20-pin connector which is used to transfer power from the battery to the control board, to control speed and direction of motors and to transfer signals from rotary encoders. Block diagram of control board is shown in Fig.2.

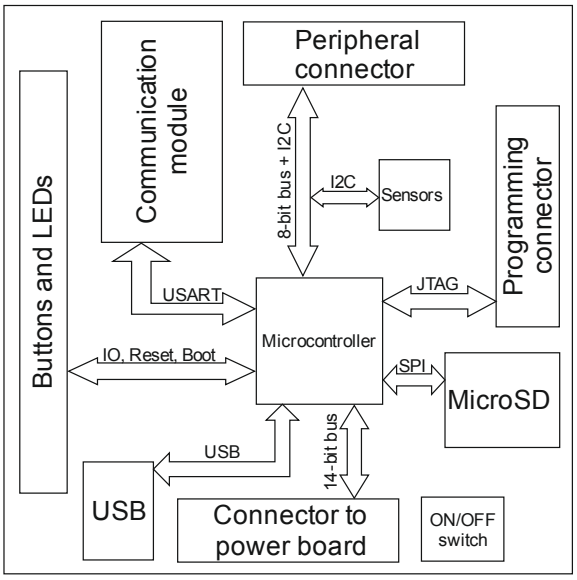


Fig. 2. Block diagram of control board.

B. *Power board*

This board is used to drive two motors, connects rotary encoders to control board and provide power from battery to control board. There are two H-bridges made up of power bipolar transistors and auxiliary logic circuits which together form a complete motor driver with direction/enable inputs. Schematic diagram of H-bridge is shown in Fig. 3.

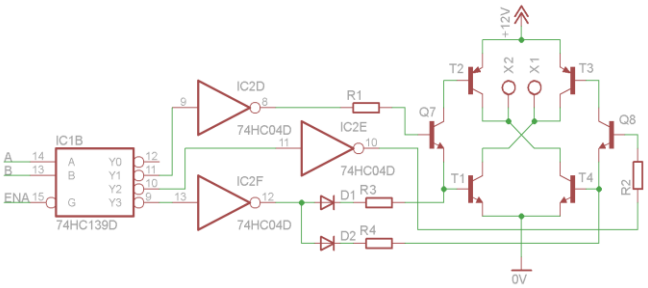


Fig. 3. Schematic diagram of H-bridge with auxiliary logic circuits.

Demultiplexer IC1 according to the input determines if the motor will turn, break or there will be no action. Inputs A and B determine action and input ENA will enable the action. Because outputs from demultiplexer are inverted, on each output is connected inverter IC2 to achieve positive logic. Demultiplexer guarantees that there won't be any prohibited states such as short caused by opening all power transistors. Table 1 shows logic states of the H-bridge depending on A, B and ENA inputs.

TABLE I
LOGIC STATES OF THE H-BRIDGE

A	B	ENA	State
X	X	1	No action
0	0	0	No action
0	1	0	Clockwise rotation
1	0	0	Counter clockwise rotation
1	1	0	Active breaking

IV. Robot's Position Calculation

Robot can calculate its own position by using magnetic encoders mounted on the wheels. Calculated position is only approximate and deviation from real position increase by travelled distance. This deviation can be decreased by using onboard gyroscope and accelerometer. When using robot as soccer player, position is determined by computer using camera placed above soccer field.

A. Encoder Sensors

Robot features two contactless magnetic rotary encoders for accurate angular measurement of each wheel. The absolute angle measurement provides instant indication of magnet's angular position with a resolution of 0.35° = 1024 positions per revolution [4]. Sensor need for its operation strong rare earth magnet with designated parameters. Magnet must be cylindrical with proper dimensions and diagonally polarized.

B. Signal Processing

Sensor can work in various modes like quadrature A/B mode and Step/Direction mode. In this case the quadrature mode is used. Output A/B signals from the sensors are directly connected to the input pins of the microcontroller. These input pins are all capable of triggering external interrupt (EXTI) so on every edge of the signal a subroutine is executed that counts impulses. Every 10 ms is then executed another subroutine that is triggered by internal timer. This subroutine calculates position from counted impulses and the result is x, y axes and φ angle of the robot.

C. Equations

From impulses received from encoders we can simply determine position and angle of the robot. These equations are discrete in time.

$$\varphi(i) = \varphi(i-1) + \frac{dr - dl}{i} \quad (3)$$

$$x(i) = x(i-1) + \frac{\cos(\varphi(i))(dr + dl)}{2b} \quad (4)$$

$$y(i) = y(i-1) + \frac{\sin(\varphi(i))(dr + dl)}{2b} \quad (5)$$

Constants i and b in equations (3), (4) and (5) depend on wheel diameter with tire (d = 62 mm), distance between the wheels (l = 67.5 mm) and number of impulses per revolution (n = 512 imp/rev). Constants can be calculated from equations (6) and (7). Description of symbols is in table 2.

$$b = \frac{n}{\pi \times d} = 26286 \quad (6)$$

$$i = l \times b = 177{,}4321 \quad (7)$$

TABLE 2
Description Of Symbols

Symbol	Description	Units
dr	Right motor counted impulses per time	imp/dt
dl	Left motor counted impulses per time	imp/dt
x	x position of robot	m
y	y position of robot	m
φ	Angle of the robot in radians	rad
i	Number of impulses per radian	imp/rad
b	Number of impulses per meter	imp/m
n	Number of impulses per revolution of the wheel	imp/r
d	Diameter of the wheel	m
l	Distance between the wheels	m
k	Gear ratio	-
τ	Motor torque	N/m
ω	Speed of the motor	rev/s
m	Mass of the robot	kg

imp = impulses, dt = time constant, mm = millimeter, rad = radian, rev = revolution of the wheel, s = second, kg = kilogram.

V. Communication

Communication with computer provides HC-05 Bluetooth module located on the control board. Bluetooth module is connected to microcontroller via two data pins transmit (Tx) and receive (Rx). Communication module provides full duplex asynchronous serial communication with baud rate up to 1382400 bits per second and with range up to 10 m. When connected to PC, module acts like virtual COM port. HC-05 also features two digital outputs to connect LEDs which serves to display status of the device.

Microcontroller uses communication interrupt when transmitting or receiving data bytes. When byte is received, it triggers interrupt handler which put byte into queue FIFO buffer. When byte is recognized as new line character, interrupt handler set semaphore that enable string recognition subroutine to be executed.

VI. Program and Operating System

Most of robot's basic functions such as calculating position of the robot, communication handling and speed control of motors are executed at the program's lowest level. These functions are executed using interrupts and provide the programmer's base tool he can work with. For higher functions such as string recognition, maze solving algorithm, and other function that consume more processing time there is real time operating system that distribute processing time for multiple programs. In this robot is used FreeRTOS (Real Time Operating System) operating system.

Most operating systems appear to allow multiple programs to execute at the same time. This is called multi-tasking. In reality, each processor core can only be running a single thread of execution at any given point in time. A part of the operating system called the scheduler is responsible for deciding which program to run when, and provides the illusion of simultaneous execution by rapidly switching between each program [5].

VII. Conclusion

Since this robot is a prototype it has few bugs we found during its construction process and we made corrections in the design to refine robot.

The next step of the work will be improvement of the motion control together with its diagnostics from integrated sensors.

In the future we plan to mount the video camera on the robot in order to make it fully autonomous and participate in the MicroMouse competition.

The most scientific contribution in the future is expected in the implementation of the playing strategies and their adaptation during the game of multiple robots.

VIII. Acknowledgement

This work has been supported by the Scientific Grant Agency of Slovak Republic under project Vega No.1/0286/11 Dynamic Hybrid Architectures of the Multiagent Network Control Systems.

References

[1] MiroSot [online]. Available: http://www.fira.net/?mid=mirosot
[2] Robosoccer.sk [online]. Available: http://www.robosoccer.sk/
[3] Micromouse (2013, February 1) [online]. Available: http://en.wikipedia.org/wiki/Micromouse
[4] AS5040, 10 Bit 360° Programmable Magnetic Rotary Encoder Data Sheet, Revision 1.8
[5] FreeRTOS [online]. Available: http://www.freertos.org/about-RTOS.html

Crowd-Sourced User-Decision Based System Call Tracing IPS – Theoretical Model

[1]*Marek Čajkovský (2nd year)*, [2]*Ivan KLIMEK (3rd year)*
Supervisor: [3] *Martin TOMÁŠEK*

Dep. of Computers and Informatics, FEI TU of Košice, Slovak Republic

[1]marek.cajkovsky@tuke.sk, [2]ivan.klimek@tuke.sk, [3]martin.tomasek@tuke.sk

***Abstract*—This paper presents first theoretical model of Intrusion Detection and Prevention System (IDPS) which combines application tracing based on system calls and user decisions process during software classification. This model allows build user profiles which are based on decisions that user have made during software classification. We introduce application reputation as a new concept in application evaluation in computer security perspective. Our concept of application reputation is based on fact that we have seen nearly all kind of malicious software (malware) since Intrusion Detection Systems (IDS) was widely deployed. Our solution thus assumes "deny any" policy as a default action for any behavior that we did not spot before. The application reputation gained during its tracing process is thus helpful tool which assists during classification of application by user. This paper further outlines several novel approaches, such as building four various databases used for software description and one profile database for describing user behavior, opposite existing solutions which mainly uses just one database for specifying malware. Presented architecture of this approach outlines predispositions to widely deployment which can be combined with elements of crowd sourcing.**

***Keywords*—crowd-sourcing, intrusion prevention system, reputation**

I. Introduction

Attacks against computer systems are still largely successful despite the plethora of intrusion prevention solutions available today. These solutions vary from basic simple scripts to complex solutions which protecting whole computer networks. Nowadays IDS are very robust, compared to their predecessors. There are several attributes which can be observed from target system. In fact the main division of IDS can be done according to deployment or according to principle how anomaly is defined.

According to deployment we can divide IDS into two main categories:

- Network-Based IDS
- Host-Based IDS

Network-based approach examines network traffic in form of packets. The most often value that is examined in host-based approach is system calls [9]. System calls allow perform privileged tasks on the system to user. This makes intuitive sense that most malicious intention an intruder will do use system calls [10]. But collecting the data is only the first step

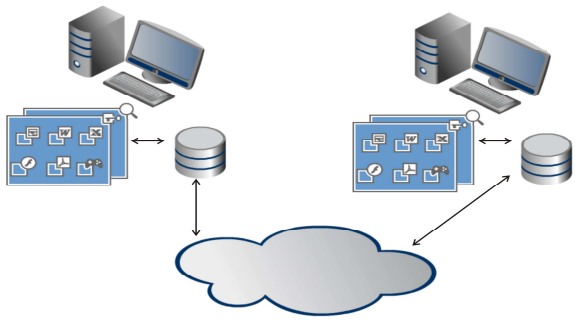


Figure 1. Client-server communication model with locally stored profiles

in IDS. Success of IDS significantly depends on procedure how those collected data are further processed. Two IDS can observe the same values on target system (collect exactly the same data), but their processing and evaluation can be completely different.

For example Kosoresow and Hofmeyr [20] described the normal process behavior by using DFA, Baliga and Lin [21] by using Kolmogorov complexity based automaton, Ghosh et al. [22] compared three different methods (equality matching algorithm, feed forward back propagation NN and Elman NN) and pinpoint their advantages and pitfalls, Smith et al. [23] introduced XFA to describe process behavior. All those papers used as their source data system calls, but used different methods to their further processing.

There are two main approaches in describing the anomaly. According to principle how anomaly is specified we can divide IDS into following two categories:

- Misuse-Based IDS
- Anomaly-Based IDS

Garcia-Teodoro et al. [3] further divides anomaly detection techniques. Misuse-based and anomaly-based approaches mainly differ in the way how the behavior is modeled and how the "normalcy" is defined. Those two approaches also specify how observed data are further processed. Further reading can be found in [1, 2, 4, 6-8].

II. Main Idea

In this section we are introducing novel approach that was not presented nor deployed so far. Proposed approach uses the rapid growth of cloud computing and high speed broadband internet as phenomena of last years, apposite to current situation in IDS, where almost every present solution is based on one of the “old" techniques or its combination and slightly differs from the original concept, proposed a few decades ago. Our system focuses to be as simple and straightforward as possible. We believe that this novel approach will combine the best features available so far.

The main idea is based on assumption that we have seen nearly all possible variants of malware since IDS concept was firstly introduced. Thus our system implicitly employs "deny any" policy for all unusual spotted behavior.

In our solution client maintains the local database of application profiles that is updated after querying cloud. Client's application is run in sand boxed environment which allows soon discovering and stopping potentially harmful applications. If profile for actually running application is missing in client's local database, the cloud is requested. We believe that average user is using only small set of applications typical to him and did not install new software every day. Thus we store this small set of profiles which are typical for user in local profiles database which is deployed on user client PC. This will save computing power on cloud that would be needed if all user application behavior was compared against database on cloud. Further it saves network traffic utilization that would be needed for particular queries. Most important is fact that it will also reduce the latency that would affect using of application. Client thus maintains significantly smaller and faster application profile database tailored for his needs. Proposed architecture is based on client-server model. Simplified architecture is shown on Fig. 1.

III. Initial Database Bootstrapping

To achieve mentioned functionality several steps are needed. As it was mentioned our proposed systems assumes "deny any" policy by default. This means that we are considering as a normal behavior only that behavior we have already seen before and we are sure that this behavior was legitimate. This implies that we need to know as much application behavior as it is possible. To achieve this task the initial database bootstrapping on cloud is crucial. We are considering this goal as crucial because there is need to collect as much application profile behavior as it is possible. Only large amount of good software (goodware) and malware can ensure that upon building application profiles we will be able decide if we already seen the behavior of application and is therefore legitimate or anomalous. Bellow is this goal outlined in eight steps:

1. Installing known malware samples (available from honey networks etc.) on a clean virtual machine instance.
2. Monitoring behavior of this virtual machine and extracting behavior patterns from running executing applications.
3. Storing observed behavior and labeling it as malicious
4. Deploying new clean virtual machine.
5. Installing binaries that are guaranteed harmless (freeware applications passed latest anti-virus tests etc.) on clean virtual machine instance.
6. Monitoring behavior of this virtual machine and extracting behavior patterns from executing applications.
7. Storing observed behavior and labeling it as normal
8. Cross correlation of observed values by comparing gathered information against each other and finding characteristics common patterns and characteristic differences.

There are also other possibilities [9] how to scan binaries for malicious behavior or finding fractions of code, such as static analysis [24]. We have chosen dynamic analysis in virtual environment, because this way is nearly similar to real situation and it is more natural way of observing true application behavior. There are several papers which deals with this problematic [12-19]

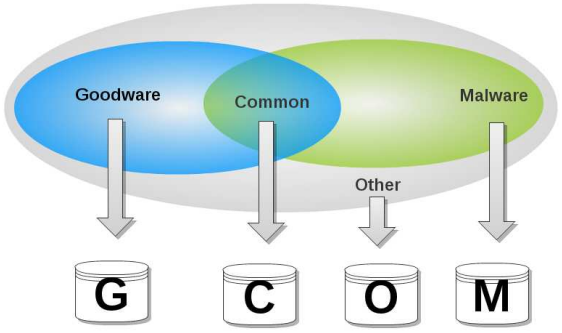


Figure 2. Extracting data during initial database bootstrapping procedure

It should be noted that build application profiles for that large amount of software is not trivial task. Today we have available inexhaustible amount of malware and goodware which is still growing. We assume using machine learning techniques as complement to formal methods since formal methods alone are not sufficient to build secure systems [5]. It is impossible observe all existing goodware or malware and extract its application profile. Moreover the full behavior of the application can only hardly be captured, and thus some other techniques which allow generalizing and unburden the human interaction are needed. Thus we are assuming using combination of various techniques as it is often presented in later years. This process should ensure that our approach will be able extract patterns that are typical for malicious and normal behavior. Fig. 2 represents the various data sources and appropriate application behavioral profiles which are generated from those data.

We propose solution where cloud in initial state creates, and later maintains four different application profiles: goodware, malware, common and other. Bellow is briefly explained main idea behind each of them:

- G - Goodware - Represents patterns that is guaranteed to be unique for harmless software.
- M - Malware - Represents patterns that is guaranteed to be unique for harmful software.
- C - Common - Represents patterns that share common signs among malware and goodware,

those patterns can be marked as an untrusted.

- - Other - Represents patterns which did not fit into three previous types, those patterns can be marked as yet unseen.

We assume that maintaining several profiles for application behavior will allow more easy, scalable and comfortable manipulation of those profiles, e.g. reclassifying profiles between goodware and malware etc. We also assume that other profile will be reducing over time, while goodware and mallware will be growing, because of nature of our solution. We believe this achieve that the knowledge-base of our solution will be easy scalable and will exhibit more accuracy results during time.

IV. Basic Working Principle

Following lines briefly explains how our system works in general.

After initial bootstrapping of cloud, client's current application behavior is compared against those bootstrapped application profiles on cloud. After this step, client's requests cloud to verify the trustworthiness of application that is about to be executed. Client then compares results from cloud. As it was mentioned we assumes "deny any" implicit policy, this means that we assume application harmless only if it's contained in goodware profile.

If application did match neither in goodware nor in malware profile, client is notified that application may be dangerous, and assumption about running is left to the user. This is crucial fact which will allow us to create user profile database, based on this fact user profile database can be modeled as following:

Application Profile + User Decision = User Profile

For example if software that is not contained in goodware and malware database appears, user receives the report from cloud in form: *"This piece of code can be: goodware in 20%, malware in 60%, common in 10%, Other in 10%"*. User after this message will face decision if he will run the software or not.

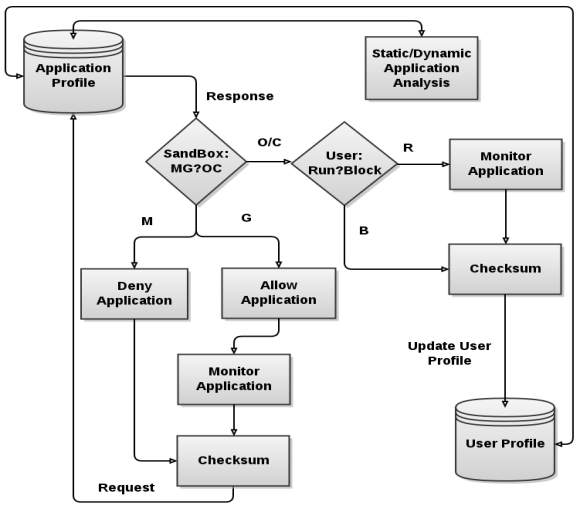


Figure 3. Simplified principle of proposed solution

User profile database can be helpful to further specify behavior of particular user of network, for example it can answer questions type: *"which user have tendency to run untrusted applications?"*, *"can be user considered as trustworthy in respect to applications that he used to run?"* or *"which user suddenly changed his behavior upon issuing untrusted application?"*

As our solution maintains several large databases (four for application profile plus user profile database) those large information have predisposition to be used further in data mining process. We believe that this approach will also allow applying further restrictions or policies per user. Applications that were not contained in goodware nor in malware yet and are executed by user are automatically send to the cloud for further analysis. This should ensure that database of profiles are always up to date. Applications that were not categorized yet (the "other" profile) are upon sending to cloud periodically submitted to further analysis (static and dynamic) and further categorized into one of profiles. This approach is shown on Fig. 3.

There are several crucial facts that need to be solved in or approach. First is building the normal and abnormal profile for anomaly specification. We are aware that building the application profile database is non trivial task, and is not possible to observe full software behavior. There are several ways how to do this. We believe that during observing applications in virtual environment we will be able to use similar methods that are used e.g. during testing in software development. This will describe as much behavior as we will need. We will focus our future research on several experiments from which we will choose those with best results.

V. Key Benefits

We assume that this novel security approach will contribute in several ways. First, our proposal solution uses four categories of data (malware, goodware, common, other) comparing to nowadays solutions which often uses just one (malware) and the other type is implicitly inferred as an negation of this one (all that is not malware is considered as goodware in misuse based approach, and all that is deviated from normal behavior is considered as anomalous behavior in behavioral based). Those four different categories can be scaled by needs of system, for example by their various combining, adding priority or changing the rules dynamically. Second, we believe that using the sand boxed environment is crucial because of software behavior can change in time, and can suddenly behaves anomaly as it is for example in various time/logic bombs. Third benefit we consider is the implicit "deny any" policy that we assume. With more restrictive security model is more difficult to compromise target system. Fourth benefit can be that the bootstrapped knowledge database can serve as more actual version of database provided by University of New Mexico. Last but not least significant benefit we consider that our approach provides overview about users in observed network. For example if user runs untrusted binaries very often she can be also considered as untrusted and further security enforcement can be applied to restrict him. This approach moves the traditional computer security as it is already known to the next level where crowd sourcing can be used.

VI. Conclusion

Nowadays security threats behave completely different than it was in the past. Malware hide its presence on target system and try to protect itself from various antivirus software. Most security threats nowadays are not intended to harm victim directly, but utilize the resources that victim provides and controls victim system remotely. Using victim computers to commit further crimes is phenomena of last years which are caused by botnets. There is another type of threats which exploit user trustworthiness and careless, those type often spy user computer and stoles private data, for example related to banks or credentials to online services. This is done by using various forms of phishing, pharming or using key loggers. It should be noted that even if intrusion detection systems have significantly evolved from main concept, proposed several decades ago, they core remain the same. Thus these solutions hardly reflect the current requirements and available technologies, solutions and knowledge available today. We believe that proposed solution which is presented in fifth chapter is answer to today state. Our approach uses the rapid growth of cloud computing and high speed internet broadband as phenomena of today. We presented more restrictive model of computer security for intrusion detection. We also rely on fact that for almost thirty years that intrusion detection systems exist we have seen all kind of behavior as well as malicious and normal. Thus our system implicitly assumes that everything that is deviated from this behavior, that we already know can be classified as suspicious

Acknowledgment

This work was supported by the Slovak Research and Development Agency under the contract No. APVV-0008-10.

References

[1] Herve Debar, Marc Dacier, and Andreas Wespi. Towards a taxonomy of intrusion-detection systems. Computer Networks, (October 1998), 1999. URL http://www.sciencedirect.com/science/article/pii/S1389128698000176.

[2] Anita K. Jones and Rrobert S. Sielken. Computer system intrusion detection: A survey. Computer Science Technical Report, pages 1-25, 2000. URL http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.24.7802\&rep=rep1\&type=pdf.

[3] P Garcia-Teodoro, J Diaz-Verdejo, G Macia-Fernandez, and E Vazquez. Anomaly-based network intrusion detection: Techniques, systems and challenges. Computers & Security, 28 (1-2):18-28, 2009. doi: 10.1016/j.cose.2008.08.003. URL http://www.sciencedirect.com/science/article/pii/S0167404808000692.

[4] Aleksandar Lazarevic, Levent Ertoz, Vipin Kumar, Aysel Ozgur, and Jaideep Srivastava. A comparative study of anomaly detection schemes in network intrusion detection.

[5] Wagner, David, and R. Dean. "Intrusion detection via static analysis." Security and Privacy, 2001. S&P 2001. Proceedings. 2001 IEEE Symposium on. IEEE, 2001.

[6] Mrutyunjaya Panda and Manas Ranjan Patra. Network intrusion detection using naive bayes. journal of computer science and network security, 7(12):258-263, 2007. URL http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.128.936&rep=rep1&type=pdf

[7] Wenke Lee and Salvatore J Stolfo. Data mining approaches for intrusion detection, volume 3. 2000. ISBN 3060296103. URL http://dl.acm.org/citation.cfm?id=382914 http://static.usenix.org/publications/library/proceedings/sec98/full_papers/lee/lee_html/lee.html.

[8] Nong Ye. A Markov Chain Model of Temporal Behavior for Anomaly Detection. (4):6-7,2000.

[9] R. Gopalakrishna, Eugene H. Spafford, and J. Vitek. Effcient intrusion detection using automaton inlining. Security and Privacy,, pages 18-31, 2005. doi: 10.1109/SP.2005.1. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1425056http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1425056

[10] Sean Peisert, Matt Bishop, Sidney Karin, and Keith Marzullo. Analysis of computer intrusions using sequences of function calls. IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING, 4(2):137-150, 2007. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4198178.

[11] Joao B D Cabrera, Lundy Lewis, and Raman K Mehra. Detection and classiffication of intrusions and faults using sequences of system calls. Acm sigmod record, 2001. URL http://dl.acm.org/citation.cfm?id=604269.

[12] Koichi Onoue, Yoshihiro Oyama, and Akinori Yonezawa. 2008. Control of system calls from outside of virtual machines. In Proceedings of the 2008 ACM symposium on Applied computing (SAC '08). ACM, New York, NY, USA, 2116-1221. DOI=10.1145/1363686.1364196 http://doi.acm.org/10.1145/1363686.1364196

[13] Monirul I. Sharif, Wenke Lee, Weidong Cui, and Andrea Lanzi. 2009. Secure in-VM monitoring using hardware virtualization. In Proceedings of the 16th ACM conference on Computer and communications security (CCS '09). ACM, New York, NY, USA, 477-487. DOI=10.1145/1653662.1653720 http://doi.acm.org/10.1145/1653662.1653720

[14] Xuxian Jiang and Xinyuan Wang. 2007. "Out-of-the-Box" monitoring of VM-based high-interaction honeypots. In Proceedings of the 10th international conference on Recent advances in intrusion detection (RAID'07), Christopher Kruegel, Richard Lippmann, and Andrew Clark (Eds.). Springer-Verlag, Berlin, Heidelberg, 198-218.

[15] Xuxian Jiang, Xinyuan Wang, and Dongyan Xu. 2007. Stealthy malware detection through vmm-based "out-of-the-box" semantic view reconstruction. In Proceedings of the 14th ACM conference on Computer and communications security (CCS '07). ACM, New York, NY, USA, 128-138. DOI=10.1145/1315245.1315262 http://doi.acm.org/10.1145/1315245.1315262

[16] Bo Li, Jianxin Li, Tianyu Wo, Chunming Hu, Liang Zhong, "A VMM-Based System Call Interposition Framework for Program Monitoring," icpads, pp.706-711, 2010 IEEE 16th International Conference on Parallel and Distributed Systems, 2010

[17] Jonas Pfoh, Christian Schneider, and Claudia Eckert. 2011. Nitro: hardware-based system call tracing for virtual machines. In Proceedings of the 6th International conference on Advances in information and computer security (IWSEC'11), Tetsu Iwata and Masakatsu Nishigaki (Eds.). Springer-Verlag, Berlin, Heidelberg, 96-112.

[18] Samuel T. King, George W. Dunlap, and Peter M. Chen. 2003. Operating system support for virtual machines. In Proceedings of the annual conference on USENIX Annual Technical Conference (ATEC '03). USENIX Association, Berkeley, CA, USA, 6-6.

[19] Cao, Ying, et al. "Osiris: A Malware Behavior Capturing System Implemented at Virtual Machine Monitor Layer." Computational Intelligence and Security (CIS), 2012 Eighth International Conference on. IEEE, 2012.

[20] Andrew P. Kosoresow and Steven A. Hofmeyr. Intrusion detection via system call traces. Software, IEEE, pages 35-42, 1997. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=605929.

[21] Priya Baliga and T. Y. Lin. Kolmogorov complexity based automata modeling for intrusion detection. Granular Computing, 2005, pages 387-392, URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1547318.

[22] Anup K. Ghosh, Aaron Schwartzbard, and Michael Schatz. Learning program behavior profiles for intrusion detection, 1999. URL http://static.usenix.org/event/detection99/full_papers/ghosh/ghosh.pdf.

[23] Randy Smith, Cristian Estan, and Somesh Jha. XFA: Faster signature matching with extended automata. In Security and Privacy, 2008. SP 2008. IEEE Symposium, pages 187-201. IEEE, 2008. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4531153.

[24] Ulrich Bayer, Christopher Kruegel, and Engin Kirda. TTAnalyze: A tool for analyzing malware. 2006. URL https://www.auto.tuwien.ac.at/$\sim$chris/research/doc/eicar06_ttanalyze.pdf.

Decision-making framework for software vendors considering transition of an application into SaaS

[1]*Adela TUŠANOVÁ (3rd year)*
Supervisor: [2]Ján PARALIČ

[1,2]Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

[1]adela.tusanova@tuke.sk, [2]jan.paralic@tuke.sk

***Abstract*— The software industry is in an evolving situation. Independent software vendors (ISVs) are facing challenges to resist not only the current economic crisis but also a fundamental shift in customer dynamic preferences relative to how they purchase and use applications. A combination of market forces is driving a growing number of organizations of all sizes to adopt Software-as-a-Service (SaaS) solutions to achieve their business objectives. SaaS solutions lead to greater success among IT vendors, opening the door to new business opportunities. This paper provides the framework to help ISVs overcome the challenges and make the best strategic decision when considering application transition to the cloud.**

***Keywords*— cloud computing, software as a service, decision-making, framework, software vendor**

I. Introduction

Sales of traditional software products and license fees decreased in recent years, while revenues of software companies have moved to selling services. These changes in the business have implications for both users and service providers. Innovation supported by services represents a competitive advantage for the company and added value for customers. According to research at the Massachusetts Institute of Technology software companies' profit from product sales decreased from 70% to less than 30% while profit from the sale of services is rapidly increasing [1]. One example is American Oracle for which profits from the sale of services overtook profits from the sale of products already around in 1997 and are still rising.

In recent years, the transition from traditional software products to services is supported by new technologies, especially cloud solutions. Cloud computing as a concept based on the provision of IT in the form of services brings also new business models, new pricing, revenue streams as well as enables achievement of new segments.

Analyst firm Gartner predicts that IT services will dramatically change business models of providers and how end users will pay for IT services. It also estimates that by 2015, 70% of existing contracts will be revised and 80% of them will result in the adoption of cloud services [2].

The results of the research should help independent suppliers of software solutions (called **I**ndependent **S**oftware **V**endors - ISVs). ISVs are mostly small and medium-sized companies that develop their own software and then sell it to the end customer. The goal is to develop the framework which will help IVSs in strategic decision to move or not to move from traditional software solutions to the cloud services (so called cloudification).

II. Proposed Framework

Proposed framework is based on the assumption that one of the key factors of company's strategic decision are financial streams – thus calculation of costs, expected revenues and return of investment are important part of the framework. Finances are not the only criteria that influence final decision of the company; therefore, there is the natural need for the multi-criteria decision making and the inclusion of other criteria that may significantly affect the company's final decision.

Proposed framework consists of seven steps (see Fig. 1), which are described in greater details in the following.

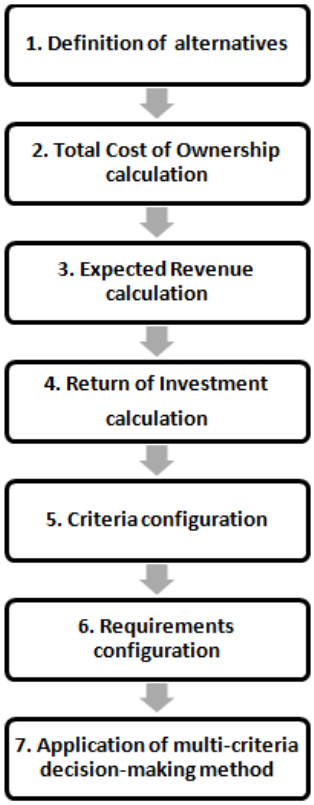


Fig. 1 Decision making framework for software vendors

A. *Definition of Alternatives*

We assume that the ISV knows what the current situation of the company is – he knows the segment in which the company operates, its products and services, its customers and competitors. It is important to consider the whole context about why the company is considering the move to the cloud and what the expected results are. Moreover, We assume that ISV has done a basic analysis of the suitability of the software transition to the cloud. Not every type of software is suitable for cloud – especially not solutions that work with sensitive data or are part of critical processes in the company.

The first step of the proposed model is the selection of alternatives which ISV considers and wants to compare. Alternatives are the different options of cloudification among which a decision maker will choose the best one. These alternatives need to be explicitly specified in the beginning of the decision process. At least two alternatives should be defined for particular workload. It is recommended that one of these alternatives should be existing solution, so it's possible to compare traditional software (i.e. continuation in current situation of the ISV) with cloud alternatives. In this framework, We take into account four possible alternatives (see Fig. 2):

A1. **S**oftware **a**s **a** **P**roduct (SaaP) - Traditional software which is sold with a standard perpetual license,

A2. **S**oftware **a**s **a** **S**ervice (SaaS) solution, where ISV considers building own datacenter, as well as providing a platform and the application itself,

A3. **S**oftware **a**s **a** **S**ervice (SaaS) solution, where infrastructure is provided by external cloud providers and ISV provides platform and application development itself,

A4. **S**oftware **a**s **a** **S**ervice (SaaS) solution, where infrastructure and platform are provided by external cloud providers and ISV is focused on application development only.

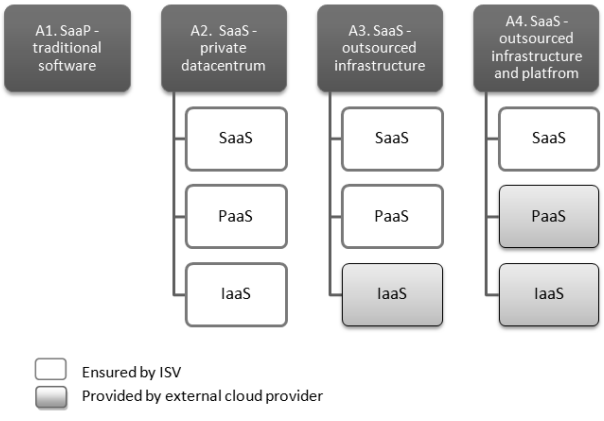


Fig. 2 Possible alternatives proposed in the framework

The selection of alternatives is an important input for further steps in the proposed framework. Calculation of total cost of ownership, expected revenue streams, return on investment and application of multi-criteria decision-making methods strongly depends on characteristics of these alternatives. To summarize the main characteristics of the first step in proposed decision framework:

Objectives: Define considered alternatives.
Inputs: Analysis of the current situation of the company.
Output: Specified considered alternatives.

B. *Total Cost of Ownership calculation*

Total **C**ost of **O**wnership (TCO) [3]**Chyba! Nenašiel sa žiaden zdroj odkazov.** is the foundation of every IT budget compilation and requires a detailed analysis. Cost is usually one of the main criteria to decide on adoption or rejection of considered alternatives. Therefore, TCO calculation is presented in the second step of the proposed framework and in next steps is considered as one of the criteria. TCO expresses the total cost of implementation and operation of particular application (e.g. planning costs, costs on design, development and testing of an application, operation and maintenance costs, costs of support and continual improvements, costs of datacenter or costs of application retirement). In the context of this framework we are calculating TCO for each alternative differently.

There are many cost analyses of traditional or SaaS software solutions, but individual cost items are different from each other or some of them are neglected. Therefore, summary of these analyzes and design of comprehensive taxonomy of the costs related to chosen alternatives is part of the framework, but the details are not presented in this paper, but are part of my dissertation work.

Objectives: Calculate total costs of considered alternatives.
Inputs: Defined considered alternatives.
Output: The total cost of each considered alternative.

C. *Expected Revenue Calculation*

One of the fundamental differences between licensed software and SaaS solution is a type of income. Companies which sell traditional software are profiting from the sale of software licenses and customer support. Normally it is a one-off sale of the product license. Resale of license comes with a new version of the product.

In contrast, SaaS model delivers recurring income, so called Recurring Revenue [4], which flow to the company on a regular basis. Customers pay for the services they actually consume, not the software itself. In the framework, we take into account the difference in revenue streams and enable the comparison of estimated income of selected alternatives.

Objectives: Calculate total revenue estimation for the considered alternatives.
Inputs: Defined considered alternatives.
Output: Estimated total revenues of the considered alternatives.

D. *Return of Investment calculation*

Return **o**f **I**nvestment (ROI) [5] is one of the most common indicators for investment decisions. This indicator takes into account all revenues, costs (including upfront costs) for a certain period of time.

Since the stream of SaaS recurring revenue and their calculation is substandard, extended and improved calculation of ROI for SaaS solution is needed. Also, there are different cost and revenue items for different alternatives, which are taken into account.

Objectives: Calculate return on investment of all alternatives under consideration.
Inputs: Defined considered alternatives, the total cost of

each considered alternative and estimated total revenues of the considered alternatives.

Output: Return on investment of all alternatives under consideration.

E. Criteria configuration

This step is necessary to define the criteria under which the company will decide. We have divided criteria extracted from published literature into four categories: business criteria (e.g. competitive advantage, knowledge of user behavior, access to new markets, time needed to delivery, avoidance of illegal copies of software, company image, quality of customer support, predictable income), financial criteria (Total Cost of Ownership, expected revenue, Return of Investment, Payback period), technical criteria (number of developed versions, scalability, software control, data security, integration, supporting tools, availability of services, technological obsolescence) and criteria of risk (cannibalization of existing solution, SLA – Service Level Agreement violence, non-compliance with legislation, failure to implement, incorrectly adjusted business model, loss of data, change management, disaster recovery, lock-in of application and data).

After defining the criteria, it is required to determine the weight of each criterion which is involved in the overall decision-making. Weights of criteria depend on the current company situation and goals –for one, it can by cost reduction and for another one expansion to new segments no matter how much it costs. Therefore, adjusting criteria weights is left to the company itself. After weighting, values to each criterion depending on the considered alternatives are assigned.

Objectives: Define the decision criteria to determine their weight and then evaluate them.

Inputs: Defined considered alternatives, the total cost of each considered alternative, estimated total revenues of the considered alternatives, return on investment of all alternatives under consideration.

Output: Configured and evaluated criteria.

F. Requirements configuration

Before applying multi-criteria decision making method itself, requirements for each criteria are defined. Requirements will help to filter out those alternatives that are not feasible in a given scenario. The requirement of the scenario is defined as the minimum or the maximum value of the selected criterion (or several criteria). Each alternative with at least one criterion, where the value is less than the minimum or greater than the maximum limit is not feasible and thus filtered as inappropriate. The alternative, which was filtered out will not be further under consideration.

Objectives: Filter out unsuitable alternatives based on defined requirements.

Inputs: Configured and evaluated criteria, requirements for individual criteria.

Output: Alternatives matching defined requirements.

G. Application of multi-criteria decision-making method

The last step is the selection of a suitable multi-criteria decision making method and its subsequent application. A suitable method is SAW (**S**imple **A**dditive **W**eighting) [5] or the method of ANP (**A**nalytic **N**etwork **P**rocess) [7], which takes into account the hierarchy of criteria and can also deal with their eventual dependency of criteria.

By applying multi-criteria decision-making method we can evaluate which alternative is the best for the company.

Objectives: Select the most appropriate alternative.

Inputs: Configured and evaluated criteria, alternatives matching defined requirements.

Output: The best alternative.

III. CONCLUSION

The SaaS market is growing rapidly in response to unprecedented business challenges. The growth of SaaS is attracting an increasing number of ISVs seeking to respond to their customers' changing business requirements. Successfully making the transition to SaaS requires more than simply re-architecting an ISV's software products. It also means a complex set of enabling technologies and service delivery capabilities in order to provide reliable, secure and scalable SaaS solutions which meet their customers' evolving needs.

When it comes to moving applications to a cloud environment, it is important to analyze all relevant factors that go into the application transition decision. Therefore we propose methodical approach based on business, technical and financial criteria. Risk is also taken into account.

In this paper, we've introduced decision-making framework for ISVs, considering application transformation to the cloud. It defines step-by-step decision-making process from alternative definition to selection of the best alternative. The framework includes calculation of total cost of ownership, expected revenues and also returns on investment.

Fields of future work include implementation of the framework, which will be available as a web application. The evaluation of the proposed framework by small and medium Slovak companies, which develop and sell their own traditional software products is needed.

ACKNOWLEDGMENT

The work presented in this paper was partially supported by the Slovak Grant Agency of Ministry of Education and Academy of Science of the Slovak Republic under grant No. 1/1147/12 (50%). The work is also the result of project implementation: Development of the Center of Information and Communication Technologies for Knowledge Systems (ITMS project code: 26220120030) supported by the Research & Development Operational Program funded by the ERDF (50%).

REFERENCES

[1] Michael A. Cusumano. 2008. The Changing Software Business: Moving from Products to Services. Computer 41, 1 (January 2008), 20-27. DOI=10.1109/MC.2008.29 http://dx.doi.org/10.1109/MC.2008.29

[2] Gartner: Predicts 2012: The IT Services Journey Toward IT Industrialization Continues to Transform the Way Service Providers and Clients Engage, http://my.gartner.com/portal/server.pt?open=512&objID=249&mode=2&PageID=864059&resId=1863714&ref=Browse

[3] COWAN, Daniel F. Total Cost of Ownership. In: Informatics for the Clinical Laboratory: A Practical Guide. Springer New York, 2005. p. 87-97.

[4] SaaS Metrics – A Guide to Measuring and Improving What Matters, http://www.forentrepreneurs.com/saas-metrics/

[5] Phillips, Jack J. Return on investment. Gulf Publ., 1997.

[6] Yoon, K. P. and C. Hwang (1995) Multiple Attribute Decision Making: An Introduc-tion, Sage Publications, London.

[7] Saaty, Thomas L. (1996). Decision Making with Dependence and Feedback: The Analytic Network Process. Pittsburgh, Pennsylvania: RWS Publications. ISBN 0-9620317-9-8.

Domain-specific Grammar Refactoring Operators

[1]*Ivan Halupka (2^{nd} year)*
Supervisor: [2]*Ján Kollár*

[1,2]Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

[1]ivan.halupka@tuke.sk, [2]jan.kollar@tuke.sk

***Abstract*—Grammar refactoring operators provide general transformations on context-free grammars, and as such are of great significance to the field of automated grammar refactoring. However, because of their generality, they do not provide efficient solutions for variety of refactoring tasks, and consequences of their application on quality attributes of a context-free grammar are not always obvious. In this paper, we present domain-specific grammar refactoring operators, which we call grammar refactoring patterns, and we also present formal language for their specification. Grammar refactoring patterns provide problem-oriented transformations for context-free grammars. They also provide some additional data, such as preconditions that grammar should fulfill in order to transformation provided by pattern be considered efficient solution, and impact of specific transformation on grammars quality attributes.**

***Keywords*—Grammar refactoring, refactoring operators, grammar refactoring patterns.**

I. Introduction

Grammarware engineering is an up-and-rising discipline in software engineering, which aims to solve many issues in grammar development, and promises an overall rise in the quality of grammars that are produced, and in the productivity of their development [1]. Grammar refactoring is a process that may occur in many fields of grammarware engineering, e.g. grammar recovery, evolution and customization [1]. In fact, it is one of five core processes occurring in grammar evolution, alongside grammar extension, restriction, error correction and recovery [6]. The problem is that, unlike program refactoring, which is well-established practice, grammar refactoring is little understood and little practised [1].

In our previous work [5], we addressed this issue by proposing an evolutionary algorithm for automated task-driven refactoring of context-free grammars. The algorithm is called mARTINICA (metrics Automated Refactoring Task-driven INcremental syntactIC Algorithm) and its key feature is that it performs refactoring on the basis of user defined refactoring task, rather then operating under some fixed objective of refactoring process. The main idea behind mARTINICA is to apply a sequence of simple refactoring operators on a chosen context-free grammar in order to produce an equivalent grammar with the desired properties. Each refactoring operator transforms arbitrary context-free grammar into equivalent context-free grammar which may have different form than original grammar.

Refactoring operators with respect to diversity of possible requirements on qualitative properties of context-free grammars provide relatively universal grammar transformations. However, because of their weak binding to actual refactoring problems, refactoring operators do not provide efficient solutions for the vast majority of domain-specific refactoring problems for which more sophisticated refactoring approaches exist. Examples of refactoring tasks for which sophisticated domain-specific approaches exist are introduction of left recursion [8] and removal of left recursion [9] in context-free grammar's production rules. In order to uniformly express solutions to these domain-specific refactoring problems and with the aim of using them in the same fashion as refactoring operators, we propose formal specification language based on patterns, which in the context of grammar refactoring we regard as problem-specific refactoring operators.

Pattern in general is a problem-solution pair in given context [2][3]. Christopher Alexander argues that each pattern can be understood as an element of reality, and as an element of language [2]. Pattern as an element of reality is a relation between specific context, certain system of forces recurring in given context and certain spatial configuration that leads to balance in a given system of forces [2]. Pattern as an element of language is an instruction, which shows how certain spatial configuration can be repeatedly used in order to balance certain system of forces wherever specific context makes it relevant [2].

As such, patterns are tools of documenting of existing, well proven design knowledge, and they support construction of systems with predictable properties and quality attributes [3]. In our view, role of patterns in the field of grammar refactoring is to preserve knowledge of language engineers about when and how to refactor context-free grammars and to support process of grammar refactoring by providing this knowledge. In order to incorporate patterns in the process of automated grammar refactoring we have coined new term - grammar refactoring patterns. Each grammar refactoring pattern describes a way in which a context-free grammar can be transformed, with preserving of language that it generates, specific situation in which this transformation is possible and consequences of this transformation on specific quality attributes of a context-free grammar. Description of situation in which transformation provided by specific pattern can be applied on specific grammar defines refactoring problem that pattern addresses. Grammar transformation provided by pattern defines solution of refactoring problem. Description of consequences of applying transformation provides context in which pattern should be used.

II. Refactoring operators

Formally, a grammar refactoring operator is a function that takes some context-free grammar $G = (N, T, R, S)$ and uses

it as a basis for creating a new grammar $G' = (N', T', R', S')$ equivalent to grammar G. This function may also require some additional arguments, known as operator parameters. We refer to each assignment of actual values to the required operator parameters of the specific grammar refactoring operator as refactoring operator instantiation, and an instance of this refactoring operator is referred to as a specific grammar refactoring operator with assigned actual values of its required operator parameters.

At this stage of development, our operator suite consists of eight refactoring operators, e.g. Unfold, Fold, Remove, Pack, Extend, Reduce, Split and Nop. The first three have been adopted from R. Lämmel's paper on grammar adaptation [4], while others have been proposed by us [5]. These operators can be applied on arbitrary context-free grammar, whose production rules are expressed in BNF notation.

Nop is operator of identical transformation, and as such it does not impose any changes on context-free grammar. Unfold replaces each occurrence of specific non-terminal within some subset of production rules with right side of production rules whose left side is this non-terminal, and in BNF notation this transformation can lead to increase in number of production rules. Fold replaces some symbol sequences on the right side of some subset of grammar's production rules with specific non-terminal, whose right side is this sequence of symbols, and as such this operator provides inverse function to unfold operator. Remove operator removes specific non-terminal and all production rules containing this non-terminal on their right or left sides from grammar, but only in case when this transformation does not impose changes on language that grammar generates. Pack replaces specific sequence of symbols within right side of certain production rule with new-non-terminal, and creates production rule whose left side is this non-terminal and whose right side is equivalent to this sequence. Extend introduces new non-terminal, creates production rule whose left side is this non-terminal and right side is some other non-terminal, and replaces all occurrences of this other non-terminal within some subset of grammar's production rules with this new non-terminal. Reduce operator removes multiple production rules with equivalent right sides, but only in case when this transformation preserves language that grammar generates. In case when there are multiple production rules whose left side is specific non-terminal, split creates new grammar in which each of such rules will have different non-terminal on its left side. More detailed description of individual refactoring operators can be found in [4][5][7].

In our refactoring approach [5], we use grammar refactoring operators as a tool for incremental grammar refactoring. We tend to keep the number of operators as small as possible, and we try to keep the refactoring operators as universal as possible. This is mainly because, as the base of refactoring operators grows, computational complexity of our automated refactoring algorithm also increases.

In this paper, we propose grammar refactoring patterns, as addition to base of refactoring operators. However, key difference between operators and patterns in this context is that growth in the number of refactoring patterns in base of refactoring operators does not have significant negative impact on calculation complexity of mARTINICA algorithm, and in many cases opposite is the true. This is caused by their domain-specific orientation and relatively narrow scope of refactoring tasks for which individual patterns are applicable.

III. Grammar refactoring patterns

In our view, each grammar refactoring pattern provides equivalent transformation on context-free grammars and in this sense concept of grammar refactoring patterns is closely related to concept of refactoring operators. However there are some key differences between grammar refactoring patterns and refactoring operators. First of all, refactoring operators provide problem-independent transformations, while grammar refactoring patterns provide problem-specific transformations. This means that refactoring operators provide general transformations, whose usage is not bound by any specific class of refactoring tasks, while grammar refactoring patterns provide domain-specific transformations, intended for tackling the issues of particular class of refactoring problems. Secondly, each of our refactoring operators can be applied on arbitrary context-free grammar, including the situation when form of particular grammar does not allow specific transformation to occur, in which case original grammar form is returned as a result of a transformation. On the other hand, each grammar refactoring pattern prescribes some specific pre-conditions that context-free grammar must fulfill in order to be transformable by particular refactoring pattern.

In our approach, each grammar refactoring pattern is represented as specification consisting of three elements, which are context, problem and solution. Problem determines situation in which transformation provided by specific pattern can occur, context describes consequences of transformation on a quality attributes of a context-free grammar, while transformation itself is specified in solution part of a pattern. In this notion of refactoring patterns, each refactoring operator is in fact refactoring pattern which lacks of explicit specification of a problem and a context. Problem part of a grammar refactoring pattern is described in the terms of grammar's quality attributes, and structural properties of grammar's production rules. Context part of a pattern is described in the terms of variations in values of grammar's quality attributes.

A. Specification of grammar refactoring patterns

For purpose of expressing grammar refactoring patterns and in order to incorporate them in refactoring process of mARTINICA algorithm, we propose language for formal specification of refactoring patterns called pLERO (pattern Language of Extended Refactoring Operators). Each refactoring pattern in pLERO is expressed using specification schema Fig. 1, which consists of pattern name and three sub-specifications which describe context, problem and a solution.

Context describes effect that grammar transformation provided by a solution has on a chosen grammar metrics, in terms of increase or decrease in their values. Specification of context in pLERO is a set metric-impact pairs, while each metric-impact pair describes impact of refactoring pattern on specific grammar metric. Purpose of context specification is to define a class of refactoring tasks to fulfillment of which can certain pattern contribute. In the case of mARTINICA this means that context defines a set of objective functions [5] for which it applies that usage of a pattern in a process of automated grammar refactoring can lead to the improvement of objective function's value.

We can interpret Fig. 2 as: Application of this pattern can lead to decrease in number of left recursive rules. Application

```
PATTERN: [Pattern name]
     CONTEXT:
          [Context specification]
     END_CONTEXT
     PROBLEM:
          [Problem specification]
     END_PROBLEM
     SOLUTION:
          [Solution specification]
     END_SOLUTION
END_PATTERN
```

Fig. 1. Schema of grammar refactoring pattern specification

$$
\begin{array}{l}
CONTEXT \\
\quad minimizes \quad countOfLeftRecursiveRules; \\
\quad maximizes \quad prod; \\
END\_CONTEXT
\end{array}
$$

Fig. 2. Example of pLERO context specificatio

of this pattern can lead to increase in number of production rules.

Problem defines structural properties that some production rules of a context-free grammar must and must not have, and also quality attributes that a context-free grammar must exhibit, in order to be transformed by instance of a refactoring pattern. By structure of grammar's production rule we mean order of specific terminal and non-terminal symbols on the right side of a production rule and occurrence of specific non-terminal on the left side of a production rule.

In order to specify this structure we have coined the term meta-structures of production rules. Each meta-structure of production rule is a structural model of specific sequence of arbitrary terminal and non-terminal symbols. Sequence of specific terminal and non-terminal symbols can be assigned to particular meta-structure of production rule only in case when this sequence exhibits structural properties prescribed by this meta-structure. To each assignment of sequence of specific terminal and non-terminal symbols to specific meta-structure of production rule we refer as to meta-structure instantiation, and to each meta-structure to which particular sequence of specific terminal and non-terminal symbols has been assigned we refer as to instance of meta-structure of production rule.

We distinguish between two types of meta-structures of production-rules, e.g. primitive and composite meta-structures. Three kinds of primitive meta-structures of production rules are meta-non-terminals, meta-terminals and meta-symbols, while meta-non-terminal corresponds to arbitrary non-terminal symbol, meta-terminal corresponds to arbitrary terminal symbol and meta-symbol corresponds to arbitrary symbol of context-free grammar, regardless of the fact is this symbol terminal or non-terminal. Instance of meta-non-terminal is a specific non-terminal symbol, instance of meta-terminal is a specific terminal symbol and instance of meta-symbol is either a specific terminal symbol or a specific non-terminal symbol of a context-free grammar. Composite meta-structures are structural models made of primitive meta-structures. We propose one kind of composite meta-structure, which is meta-closure. Each meta-closure is a sequence of repeating primitive meta-structures, while instance of particular meta-closure is specific sequence of terminal and non-terminal symbols. Two meta-structure instances can be compared only if they were instantiated on the basis of a same meta-structure, and they are equivalent only when they refer to the same sequence of a specific terminal and non-terminal symbols, otherwise they are not equivalent.

Structure of right side of context-free grammar's arbitrary production rule can be described by a specific sequence of meta-structures, and each production rule can be viewed as sequence of meta-structure instances on its right side and an instance of meta-non-terminal on its left side. Specification of a problem in pLERO consists of a four parts, which are declarations, positive-match, negative-match and forces.

In declarations part all meta-structure instances and composite meta-structures of grammar production rules are specified. In meta-structure instances specification we always assume that two or more different, but comparable meta-structure instances are always not equivalent. This means that if we want to allow situation where more meta-structure instances specify same sequence of terminal and non-terminal symbols, we have also to specify the number of composite meta-structures that is equivalent to the count of such meta-structure instances, and assign each meta-structure instance to different meta-structure.

$$
\begin{array}{l}
DECLARATIONS: \\
\quad Nonterminal1 : NONTERMINAL; \\
\quad ArbitrarySequence1 : \\
\quad\quad CLOSURE('ANY\_SYMBOL', 0); \\
\quad ArbitrarySequence2 : \\
\quad\quad CLOSURE('ANY\_SYMBOL', 0); \\
\quad ArbitrarySequence3 : \\
\quad\quad CLOSURE('ANY\_SYMBOL', 0); \\
\quad ArbitrarySequence4 : \\
\quad\quad CLOSURE('ANY\_SYMBOL', 0); \\
END\_DECLARATIONS
\end{array}
$$

Fig. 3. Example of pLERO declarations specification

We can interpret Fig. 3 as: One meta-non-terminal instance called Nonterminal1, Declaration of four composite meta-structures which can be matched against any number of any grammar symbols, including zero, and declaration of one instance of each composite meta-structure.

In positive-match part of pLERO problem specification, structural properties that some subset of grammar's production rules must exhibit are defined. Positive-match is defined as a set of production meta-rules. Each production meta-rule defines a structure of one production rule, and it consists of label, left side and right side of production meta-rule. Label is unambiguous identifier of production meta-rule and enables us to manipulate with whole grammar's production rule whose structure is represented by given meta-rule, while this manipulation occurs in solution part of a pLERO specification. Left side of production meta-rule is some meta-non-terminal and right side of production meta-rule is some sequence of meta-structure instances.

In order to some grammar exhibit required structural prop-

erties it must contain production rules whose structural properties match with each of production meta-rules specified in positive-match. To matching of production rule to production meta-rule we refer as to production rule labeling. Each grammar's production rule can be labeled with at most one production meta-rule, and each production-meta rule can be used for labeling at most one production rule. This however can lead to two kinds of non-determinism, non-determinism in selecting of the production rule which will be labeled, and non-determinism in selecting of production meta-rule by which production rule should be labeled. In case when there are multiple production rules which match one production meta-rule and in case when there are multiple production meta-rules to which one production rules matches, sophisticated strategy for conflict resolution is required. Specific context-free grammar exhibits required structural properties only in case when all production meta-rules have been used for labeling, but in this case not all production rules have to be labeled.

$$
\begin{array}{l}
POSITIVE\_MATCH: \\
\quad [Rule1]Nonterminal1 :: \\
\quad\quad Nonterminal1 \;\; ArbitrarySequence1; \\
\quad [Rule2]Nonterminal1 :: \\
\quad\quad ArbitrarySequence2; \\
END\_POSITIVE\_MATCH
\end{array}
$$

Fig. 4. Example of pLERO positive-match specification

We can interpret Fig. 4 as: Grammar must contain at least two production rules whose left side is a same non-terminal. Right side of one of these rules must also start with this non-terminal followed by arbitrary sequence of symbols, while right side of other rule is an arbitrary sequence of symbols. Rule1, and Rule2 are labels of corresponding production meta-rules.

In negative-match part of pLERO problem specification, structural properties that some subset of grammar's production rules must not exhibit are defined. Negative-match is defined similarly as positive-match, and it represented by a set production meta-rules, however in this case all production meta-rules are without labels. In order to grammar exhibit structural properties that prevent transformation provided by a pattern solution, similarly as in positive-match all production meta-rules in negative-match must be used for labeling.

Labeling of production rules with production meta-rules specified in positive-match, and labeling of production rules with production meta-rules specified in negative-match are two separate but interlinked processes. This means that some production rule can be labeled with one production meta-rule of positive-match and in the same time this production rule can be labeled with one production meta-rule of negative-match. However some meta-structure instance that occur in arbitrary production meta-rule of positive-match and same meta-structure instance occurring in arbitrary production meta-rule of negative-match represents in both cases same sequence of same symbols.

We can interpret Fig. 5 as: Grammar does not contain two rules whose left side is instance Nonterminal1 and whose one right side starts with this instance and other right side contains this instance.

$$
\begin{array}{l}
NEGATIVE\_MATCH: \\
\quad Nonterminal1 :: \\
\quad\quad Nonterminal1 \;\; ArbitrarySequence1; \\
\quad Nonterminal1 :: \\
\quad\quad ArbitrararySequence3 \;\; Nonterminal1 \\
\quad\quad ArbitrarySequence4; \\
END\_NEGATIVE\_MATCH
\end{array}
$$

Fig. 5. Example of pLERO negative-match specification

In forces part of pLERO problem specification additional quality attributes that grammar must posses are defined. This quality attributes are expressed using relational expressions, whose variables are grammar metrics, and logic operators between this expressions. In this part of specification meta-structure instances can also be used, as arguments of grammar metrics. From a global point of view pLERO specification of forces is one logic expression, which if evaluated as true, indicates that grammar possesses required quality attributes, and if evaluated as false, shows that grammar does not exhibit required quality attributes.

$$
\begin{array}{l}
FORCES: \\
\quad RulesLeftSide(Nonterminal1) = 2 \\
END\_FORCES
\end{array}
$$

Fig. 6. Example of pLERO forces specification

We can interpret Fig. 6 as: Grammar contains exactly two production rules whose left side is non-terminal matched against meta-non-terminal instance Nonterminal1.

Solution in pLERO provides transformation on context-free grammar, in case that this grammar possesses structural properties defined in positive-match of problem specification, exhibits quality attributes defined in forces specification and does not posses structural properties defined in negative-match of problem specification. This transformation can be only related to meta-structure instances used in positive-match specification, and production rules which have been labeled. To specify this transformation we use relatively simple imperative language, which manipulates with meta-structures and production meta-rule labels as with constants, and allows execution of some operations on context-free grammar. Since transformation provided by solution is in fact refactoring operator refactoring operator, here we will not discuss it in detail.

We can interpret Fig. 7 as: Introduce new non-terminal of grammar and match it against meta-non-terminal instance Nonterminal2, remove productions with labels Rule1 and Rule2 and introduce three new production rules. Along with Fig. 2, Fig. 3, Fig. 4, Fig. 5 and Fig. 6 this solution specifies a method of removing left-recursion in case when one grammar's production rule is left recursive and other is not.

IV. Conclusion

In this paper we presented formal specification language for preserving of knowledge of language engineers. Our ap-

$$
\begin{aligned}
&SOLUTION: \\
&\quad INTRODUCE\_NONTERMINAL \\
&\qquad (Nonterminal2); \\
&\quad REMOVE\_PRODUCTION(Rule1); \\
&\quad REMOVE\_PRODUCTION(Rule2); \\
&\quad INTRODUCE\_PRODUCTION([New1] \\
&\qquad Nonterminal1 :: ArbitrarySequence2 \\
&\qquad Nonterminal2); \\
&\quad INTRODUCE\_PRODUCTION([New2] \\
&\qquad Nonterminal2 :: ArbitrarySequence1 \\
&\qquad Nonterminal2); \\
&\quad INTRODUCE\_PRODUCTION([New3] \\
&\qquad Nonterminal2 :: EMPTY\_SYMBOL); \\
&END\_SOLUTION
\end{aligned}
$$

Fig. 7. Example of pLERO solution specification

proach is based on patterns, which are well-proven method of knowledge preservation in other application domains, such as software architectures [3]. Main advantages of this approach are its relative simplicity, universality and suitability for automated application. Our specification language is however still in development, and disadvantage of its current version is that one production meta-rule always corresponds to exactly one production rule, which means that each pattern provides transformation on production rules whose count is limited by the size of positive-match part of pLERO problem specification.

In the future, we would like to focus on resolving this issue, so each pattern can specify structure of grammar's production rules whose count is not limited by number of production meta-rules used for pattern specification. We would also like to focus on incorporating grammar refactoring patterns in the automated refactoring process of mARTINICA algorithm.

Acknowledgment

Research described in the paper was supervised by prof. Ing. Ján Kollár, CSc., FEI TUKE in Košice and supported by project VEGA 1/0341/13 Principles and methods of automated abstraction of computer languages and software development based on the semantic enrichment caused by communication.

References

[1] P. Klint ,R. Lämmel, C. Verhoef, "Toward an engineering discipline for grammarware", *ACM Transactions on Software Engineering Methodology*, vol. 14, no. 3, pp. 331 - 380, 2005.

[2] C. Alexander, *The Timeless Way of Building*. New York, USA: Oxford University Press, 1979.

[3] F. Buschmann, R. Meunier, H. Rohnert, P. Sommerlad, M. Stal, *Pattern-Oriented Software Architecture Volume 1: A System of Patterns*. New York, USA: John Wiley & Sons, 1996.

[4] R. Lämmel, "Grammar Adaptation", in *Proceedings of the International Symposium of Formal Methods Europe on Formal Methods for Increasing Software Productivity*, pp. 550 - 570, 2001.

[5] I. Halupka, J. Kollár, E. Pietriková, "A Task-driven Grammar Refactoring Algorithm", *Acta Polytechnica*, vol. 52, no. 5, pp. 51 - 57, 2012.

[6] T.L. Alves, J. Visser, "A Case Study in Grammar Engineering", in *Proceedings of SLE'2008*, pp. 285 - 304, 2008.

[7] R. Lämmel, V. Zaytsev, "An Introduction to Grammar Convergence", in *Proceedings of the 7th International Conference on Integrated Formal Methods*, pp. 246 - 260, 2009.

[8] N.A. Kraft, E.B. Duffy, B.A. Malloy, "Grammar Recovery from Parse Trees and Metrics-Guided Grammar Refactoring", *Software Engineering*, vol. 35, no. 6, pp. 780 - 794, 2009.

[9] K.C. Louden, *Compiler Construction: Principles and Practice*. Boston, USA: PWS Publishing, 1997.

Dynamic processes described in Ludics

[1]*Pavol MACKO (3^{rd} year)*
Supervisor: [2]*Valerie Novitzká*

[1,2]Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

[1]pavol.macko@tuke.sk, [2]valerie.novitzka@tuke.sk

***Abstract*—This paper deals with ludics theory and modeling real dynamic world by means ludics. First step is capture dynamics of real world to formulas in linear logic, with which it is closely related ludics theory. It is needed to work with linear formula; construct a proof and reduced proof. Then we can describe actions and construct designs, but designs correspond to strategies in ludics, which are consist from actions too. In this paper we focused to actions and their description by strategies in ludics.**

***Keywords*—Ludics, Linear logic, actions, strategies, chronicles, daimon**

I. Introduction

In our previous works [1], [3], [12], [11] we showed, that linear logic [4], [5] is very strong logic for describe dynamics in world and in informatics too. Now we can better modeling dynamics processes in linear logic with conection with theory of ludics. The theory of ludics claims also that proof do not manipulate formulas, but their location, the address where they are stored, what we showe in [2]. All experiments are described in linear logic, because the theory of ludics and linear logic are closely related [9].

II. Actions

The triplet (ϵ, ξ, I) is called an *action*[9]. As we have seen, ϵ indicates polarity; ξ is an address (the address of a formula) and I a set of indices, the relative addresses of the immediate relative subformulas we are considering. ξ is called focus of the action, while I is called ramification [7].

III. Ludics

Ludics can be sum up as a interaction theory. It appears in the work of J.-Y. Girard [6] as the issue of several changes of paradigms in Proof Theory: from provability to computation, then from computation to interaction.

The first change of paradigm arises with the intuitionnistic logic, while the second is due to the development of linear logic. Continuing the new approachs of Linear Logic: a geometrical point of view of proofs [4], [10]; an internal approach of dynamics, Ludics focalizes on the interaction [9]. The objects of the Ludics are no more proofs but instead incomplete proofs, attempts of proofs. So a rule called damon is available in order to symbolize the giving up in a proof search or a pending lemma. These objects play the role of a proof architecture. Only what is needed for the interaction is kept. This has been made possible by means of the hypersequentialized linear logic introduced by J-M. Andreoli after he has discovered the polarity of formulas. Moreover, this work within polarized objects world allowed to create a link [7] between Ludics and recent works in Game Semantics which share similar motivations. So the Game Theory is a good metaphor for a first approach of Ludics, and it is the point of view that from now we shall often follow in this text. Here you find a presentation of the theory in a very simplied version, but we recommand the source texts [6] to the reader concerned with more details on the mathematical notions and rich concepts of Ludics; we also recommand the reading of this introduction [5].

A. The objects of Ludics

The central object of Ludics is the design. By means of the metaphor of Games, a design can be understood as a strategy, i.e. as a set of plays (chronicles) ending by answers of Player against the moves planned by Opposant. The plays are alternated sequences of moves (actions). The moves are defined as a 3-uplet constitued by: firstly a polarity (positive polarity for a move of Player or negative polarity for a move of Opposant), secondly a locus (a fixed position) from which the move is anchored, and at last a finite number of positions reachable in one step (ramification). A unusual positive move is also possible: (the daimon) [5].

In Ludics, the positions are addresses, loci incoded by means of a finite sequence of integers (often noted ξ, ρ, σ ...).

The starting positions (forks) are denoted $\Gamma \vdash \Delta$; where Γ and Δ are finite sets of loci such that Γ is either the empty set or a singleton one. When an element belongs to Γ, every play then starts on this element by means of an Opposant move (and the fork is said negative), else Player starts on an element of its choice taken in Δ (and the fork is said positive).

For the hypersequentialized linear logic point of view, a design can be seen as a figure of a proof in this sequent calculus with some particularities: first we can use the daimon rule, for giving up the proof search. Finally we don't work with formulas but with addresses, and we just need two rules (the negative and positive ones) for representing the usual logical rules [9], [8].

Every usual logic connectives has not its own, only two rules are suficient for subsuming these rules.

A design is a tree of forks $\Gamma \vdash \Delta$, built by means of these three rules:

Daimon
Daimon expresses stop action.

$$\frac{}{\vdash \Delta}\ \dagger$$

Positive rules
Assume I is a set of indices (called ramification [6]), and for $i \in I$ the Γ_i are pairwise disjoint and included in Γ. One can apply the following rule (finite, one premise for each $i \in I$):

$$\frac{\ldots \quad \xi.i \vdash \Delta_i \quad \ldots}{\vdash \xi, \Delta}\ (\xi, I)$$

Negative rules
Assume N is a set of ramifications (called the directory of the rule), and for all $I \in N$, $\Lambda_I \subseteq \Lambda$. One can apply the following rule (possibly infinite, one premise for each $I \in N$):

$$\frac{\ldots \quad \xi.I \vdash \Delta_I \quad \ldots}{\xi \vdash \Delta}\ (\xi, N)$$

The notation $\xi.I$ we have used is short for $\{\xi i, i \in I\}$.

IV. Example - login process

We have a client server network communication and login process between client and server. One logs on by his ID, after entering ID he has the freedom to choose whether they want to log in as a user or as an administrator (external nedeterminism). If he wants to log in as an administrator, then from the server can occur to two requirements for correct registration. Server choose one of the two options for correct login. (internal nedeterminism). The first possibility is that the server will require to correctly login only a password of administrator. The second possibility is that the server will require to correctly login: password of administrator and verification code.

From these facts we can formulated a formula of linear logic.

$$ID \multimap (user\_pass \ \& \ (admin\_pass \oplus (admin\_pass \otimes code)))$$

The proof of this formula of linear logic:

$$\dfrac{\dfrac{}{ID \vdash user\_pass} \quad \dfrac{\dfrac{}{ID \vdash admin\_pass}}{ID \vdash admin\_pass \oplus (admin\_pass \otimes code)}\ (\oplus_{-r_1})}{ID \vdash (user\_pass \,\&\, (admin\_pass \oplus (admin\_pass \otimes code)))}\ (\&_{-r})$$
$$\dfrac{}{ID \multimap (user\_pass \,\&\, (admin\_pass \oplus (admin\_pass \otimes code)))}\ (\multimap_{-r})$$

The second part of the proof, because fomula contains internal nedeterminism:

$$\dfrac{\dfrac{}{ID \vdash user\_pass} \quad \dfrac{\dfrac{\dfrac{}{ID \vdash admin\_pass} \quad \dfrac{}{ID \vdash code}}{ID \vdash admin\_pass \otimes code}\ (\otimes_{-r})}{ID \vdash admin\_pass \oplus (admin\_pass \otimes code)}\ (\oplus_{-r_1})}{ID \vdash (user\_pass \,\&\, (admin\_pass \oplus (admin\_pass \otimes code)))}\ (\&_{-r})$$
$$\dfrac{}{ID \multimap (user\_pass \,\&\, (admin\_pass \oplus (admin\_pass \otimes code)))}\ (\multimap_{-r})$$

From this proof we can construct a design, where we do not manipulate with formulas, but with their location, the address where they are stored [9]:

$$\dfrac{\dfrac{}{\xi.0.1 \vdash \xi.0.2} \quad \dfrac{\dfrac{}{\xi.0.1 \vdash \xi.0.3.1}}{\xi.0.1 \vdash \xi.0.3}\ (+, \xi.0.3, \{1\})}{\xi.0 \vdash}\ (-, \xi.0, \{\{1,2\}, \{1,3\}\})$$
$$\dfrac{}{\vdash \xi}\ (+, \xi, \{0\})$$

$$\dfrac{\dfrac{}{\xi.0.1 \vdash \xi.0.2} \quad \dfrac{\dfrac{}{\xi.0.1 \vdash \xi.0.3.2}}{\xi.0.1 \vdash \xi.0.3}\ (+, \xi.0.3, \{2\})}{\xi.0 \vdash}\ (-, \xi.0, \{\{1,2\}, \{1,3\}\})$$
$$\dfrac{}{\vdash \xi}\ (+, \xi, \{0\})$$

Each step in design is labeled by appropriate action by means rules in [9].

We are going to represent by means of designs klient's and server's strategies in Ludics. We arbitrarily start the interaction at locus ξ.

Three strategies of login in Ludics:
The three strategies begin in a same manner: A man enters his ID; represesented by a positive action

$$(+, \xi, \{0\}).$$

Server is ready to receive ID and then receive user password; represented by a negative action

$$(-, \xi.0, \{1, 2\})$$

or server is ready to receive ID and then receive administrator's login data; represented by a negative action

$$(-, \xi.0, \{1, 3\}).$$

In the first strategy, it ends with success login as user by means ID and password; in ludics, we will say that he plays †; playing the daimon in a strategy allow us to attest that the exchange of resource correctly ends. This strategy is described by means following tree:

$$\dfrac{\dfrac{}{\xi.0.1 \vdash \xi.0.2}\ \dagger \quad \dfrac{}{\xi.0.1 \vdash \xi.0.3}}{\xi.0 \vdash}$$
$$\dfrac{}{\vdash \xi}$$

In the second strategy a man choose login as administrator and server choose login by means password only; it is represented by a positive action

$$(+, \xi.0.3, \{1\})$$

Second strategy is described by means following tree:

$$\dfrac{\dfrac{}{\xi.0.1 \vdash \xi.0.2} \quad \dfrac{\dfrac{}{\xi.0.1 \vdash \xi.0.3.1}\dagger}{\xi.0.1 \vdash \xi.0.3}}{\dfrac{\xi.0 \vdash}{\vdash \xi}}$$

In the third strategy a man choose login as administrator and server choose login by means password and verification code; represented by a positive action

$$(+, \xi.0.3, \{2\}).$$

Third strategy is described by means following tree:

$$\dfrac{\dfrac{}{\xi.0.1 \vdash \xi.0.2} \quad \dfrac{\dfrac{}{\xi.0.1 \vdash \xi.0.3.2}\dagger}{\xi.0.1 \vdash \xi.0.3}}{\dfrac{\xi.0 \vdash}{\vdash \xi}}$$

V. The interaction

The designs are built on the model of proofs without cut. The underlying signification of the cut is the composition of morphisms or strategies. In Ludics, it is concretely translated by a coincidence of two loci in dual position in the bases of two designs. We can cut for example a design of base $\sigma \vdash \xi$ and a design of base $\xi \vdash \rho$, so forming a cut-net of base $\sigma \vdash \rho$.

The interaction is obtained by means of the cut; it creates a dynamics of rewriting of the cut-net ; at the end the process fails or we obtain a design with the same base as the starting cut-net. In the next goal we can describe actions through action semantics [13].

VI. Conclusion

In this contributions we presented how we can capture dynamics processes by ludics theory. All sample we are described in linear logic, because the theory of ludics and linear logic are closely related. First step is construct a formula in linear logic and a proof too. Then by means rules we can construct desing, which is base of construction strategies in ludics.

Next goal of our research are interactions and modeling interactions in ludics and develop a theory to describe adding address in proofs and actions to better captured real processes. (quantification of resources in linear logic proof).

References

[1] P. Macko, "How to apply recursion and corecursion in mathematical theory of programming," SCYR 2012 : proceedings from conference: 12th Scientific Conference of Young Researchers: May 15th, 2012, Herany, Slovakia. - Koice: TU, 2012, pp. 151–153, Herany, Slovakia. ISBN 978-80-553-0943-9.

[2] P. Macko and V. Novitzká and V. Slodičák, "The Rle of Designs in Linear Logic ,"Electrical Engineering and Informatics 3 : proceeding of the Faculty of Electrical Engineering and Informatics of the Technical University of Koice. - Koice : FEI TU, 2012 S. 620-623. - ISBN 978-80-553-0890-6.

[3] V. Slodičák and P. Macko, "The rle of linear logic in coalgebraical approach of computing," J. of Information and n. . Organizational Sciences. Vol. 35, Eds., 2011, pp. 197–213, iSBN 0-13-120486-6.

[4] J. Girard, "Linear logic," *Theoretical Computer Science*, vol. Vol. 50, no. No. 1, pp. 1–102, 1987, iSSN 0304-3975.

[5] J. Girard, "From foundations to ludics," *Bulletin of Symbolic Logic*, 2003, vol. Vol. 9, no. No. 2, pp. 131-168.

[6] J. Y. Girard, *Locus Solum: From the rules of logic to the logic of rules*, 2001, vol. Vol. 11 Issue 3, pp. 301 – 506.

[7] C. Faggian and M. Hyland, "Designs, disputes and strategies," S.-V. B. Heidelberg, Ed., 2002, pp. 442457.

[8] M. Basaldella and C. Faggian, "Ludics with repetitions (exponentials, interactive types and completeness)," 24th Annual IEEE Symposium on Logic In Computer Science, Ed., 2009.

[9] C. Faggian, "On the dynamics of ludics, A study of interaction," Thése de doctorat, 2002.

[10] J. Y. Girard, "Proofs and Types," *Cambridge University Press*, pp. 1–175, 2003.

[11] V. Slodičák and P. Macko, "New approaches in functional programming using algebras and coalgebras," in *European Joint Conferrences on Theory and Practise of Software - ETAPS 2011*. Universität des Saarlandes, Saarbrücken, Germany, March 2011, pp. 13–23, iSBN 978-963-284-188-5.

[12] V. Slodičák and P. Macko, "How to apply linear logic in coalgebraical approach of computing," C. . proceedings of the 22nd Central European Conference on Information and I. Systems, Eds., September 21st-23rd 2011, pp. 13–23, varadin: University of Zagreb, Croatia. ISSN 1847-2001.

[13] V. Slodičák and V. Novitzká, "Principles of Action Semantics for Functional Programming Languages," In: Studia Universitatis Babes-Bolyai Series Informatica. Vol. 57, no. 1 (2012), p. 35-47., ISSN 2065-9601

Dynamics of Slovak Company Network

[1]*Martin REPKA (4th year)*
Supervisor: [2]Ján PARALIČ

[1,2]Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

[1]martin.repka@tuke.sk, [2]jan.paralic@tuke.sk

***Abstract*— This paper deals with social network analysis of specific type of social networks – company networks. Main part is targeted to a brief description of a method for projection of Slovak company network onto company to company network in order to transform direct and indirect connections into one type of connections in the final network. Proposed type of projection enhances the projected information and makes it easier to view and analyze the structure characteristics in the network. This method proposes utilization of temporal data in Company Networks. Next part is targeted to analysis of time snapshots of projected network in order to get view of dynamic characteristics in this network.**

***Keywords*—Company Network, Network Analysis, Social Networks, Dynamics Network Analysis, Network Projections**

I. Introduction

Company Network (CN) is such a social network, where actors are represented by companies and connections are representing interactions between them. CN is little bit different from other kinds of social networks, because the actors are not behaving like common actors in social network. Dynamics of CNs is slower and they do not tend to create connections above certain boundaries (local business) because higher trust is needed than in other social networks.

Because of this nature of Company Networks, they are not very dense, but they usually incorporate local clusters, where density is much higher. This implies that CN will be characterized by significant local structures. Detection of such local clusters containing these structures can be used to simplify and speed up the process of other subsequent network analyses [1].

Company Networks also contain a lot of nonstructural data as compositional data and temporal data. Which allows us bring to front many improvements and modifications of classic social network analyses.

The paper is structured as follows. Second section describes Slovak Company Network (SCN) and its origin, deals also with the motivation of our research and explains certain features of SCN. Next it sets ground for our approach of CN projection with temporal data. Third chapter deals with dynamics of SCN and its projected version, analyzing especially the correlation between characteristics of these networks. In the fourth chapter a proposal of future work on dynamics analysis of local structures in Slovak company network is described.

II. Slovak Company Network

A. Motivation

Slovak company network (SCN) was built and continuously updated e.g. by ITLIS project[1]. Authors of Itlis aim to save time and costs of their users in finding and processing distributed information. Visualization of this network provides quick look at the direct relationships between people and companies (including companies themselves). SCN summarizes all legal relationships between persons and legal persons. All these legal acts are registered in insertion of particular District Court. On the page of ITLIS project[1] user can find transparent search and visualization tool that lets him/her interactively explore the relationship between the entities. It provides also useful analytical tools to help users find hidden relationships and derived information in the original data. This network could be useful mainly to lawyers, investigative journalists, financial analysts, loan providers and risk managers [2].

In this paper we will review dynamics of SCN and we will look into temporal features of SCN. To know the origin of temporal data and evolution of SCN is very essential in processing of SCN, e.g. we utilized these temporal features in our work [3].

B. Origin of Data and its Description

Data of SCN originate in publicly accessed sources as Business Register of the Slovak Republic (ORSR) [3], which was hardly collected, aggregated, analyzed and integrated within project ITLIS.eu into one social network. As a result of these efforts large network was created, featuring complete data model, which integrates structural, compositional and temporal (historical) data [5].

SCN is a complex network similar to two-mode networks. It distinguishes three types of actors, as well as three types of connections between actors. Actors represent persons and legal entities (organizations). All organization actors (C) must have its own insertion at District Court. Personal actor (P) can be any person involved in business; all persons involved within boards of companies must have records in insertion of particular company at district court. Agents are actors, who have no insertion in business register, but are mentioned in insertions of other companies. Although it cannot be decided if somebody is a person or legal person in this case,

[1] http://www.itlis.eu

but as a matter of fact it has to be one of them. Usually they are legal persons from other countries.

CN differs from classical two-mode network in the following characteristics. Firstly, it is the presence of agents as third type of actors, but for now we can omit them. So we consider two types of actors: persons (P) and organizations (companies) (C). Secondly, beside the existence of connections between P and C (relationship of affiliation of P to C) there are also direct connections between C and C, which is not common in classic two-mode affiliation networks. So from this point of view we distinguish in the following *direct connection* (connection of C to C) and *affiliate connection* (connection of P to C).

Another fact is that between two actors (P, C) there can be multiple connections at the same time, e.g. actor P can be both "managing director" and "limited partner" in the company (C) at the same time. [2].

III. Dynamics in SCN

A. Structural and Temporal features of SCN

As we mentioned before SCN features various connections, from the point of "one mode of network" we can identify connections between two companies (a direct connections of certain kind, e.g. limited partnership, ownerships, foundation). SCN features also two-mode network, where connections between actors (as physical person) and companies (legal persons) are present. These connections will be referred later as indirect (intermediated) connections in SCN.

In SCN we have full history of all relationships (connections) between actors. Every connection has its validity dates: "*valid from*" and "*valid to*". This feature of data allows us considering of aging of all connections and create a better projection of a network in its "*current state*" as well as to perform data analysis from *historical* point of view, like analyzing the dynamics of a network or growth of a network over time.

This fact allows us to see SCN as network with historical data, which allows us to examine its dynamics and creates space for snapshot (time) projections.

Interesting fact is that many techniques of projection do not consider *temporal* data. E.g. if we have two companies C1, C2 and a person P1. Person P1 had relationship (for instance "Limited partnership") with C1 for some period of time in the past.

Then P1 left the company (i.e. the respective connection broke up) and after some time P1 entered company C2. So we have two connections: from P1 to C1 and from P1 to C2. But these connections were *valid in different time periods*. Due to this fact we can assume that connection "through" P1 has none or only very limited effect on the relationship between companies C1 and C2.

Classic projection would assume a connection between companies, but using our improved method of projection we consider only "*overlapping*" connections, *weighted by the length of time interval, when they existed concurrently*. We believe that only connections from person to companies existing simultaneously can project some relationship between companies.

In the following section we will briefly describe our approach of transformation two-mode company network onto one-mode network using temporal data and then we will look closer into snapshot (time) projections and their characteristics [6].

Because of nature of Company Networks (CN), they have specific feature. They are not very dense, but they have local

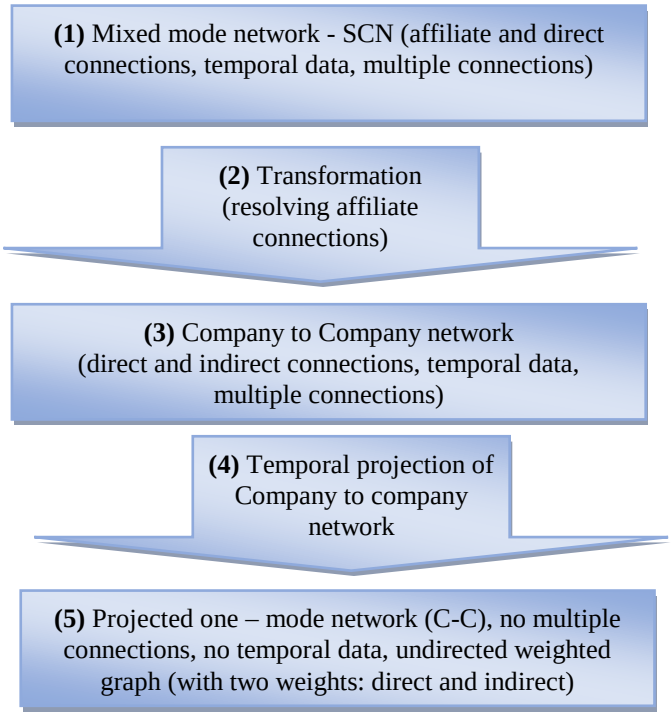


Fig. 1. Process of SCN projection on one mode company to company network.

clusters, where density is much higher. This implies that SCN will be characterized by significant local structures. These structures can be detected as local clusters and we can simplify or speed up the process of determining local structure and local role equivalence definition for this type of network. As we already succeeded to detect local structures (see [6], [7], [8]) the next interesting step is to analyze evolution of local structures in time, but we will not deal with this structural feature in details within the content of this paper.

B. Projections of Slovak company network

In work [6] we have presented several approaches to temporal projections of Company networks. In recent time we have improved and simplified whole process as follows (see Fig.1):

Firstly **(1)** our SCN in "mixed" mode features affiliate P-C and direct C-C connections (all with temporal data). In this state the network can be modeled as directed multigraph, because the connections have directions (P->C, C->C) and multiple connections between two actors are present.

Secondly **(2)** runs the process of transformation of affiliate (C->P->C) connections into "indirect" C-C connections. In this step weighting can be applied, based on the selected method of projections (see [6]).

Now **(3)** the Company network is free of person actors (P), but still contains multiple connections and temporal data.

Next **(4)** finally temporal projection takes place (aging of connections and weighting).

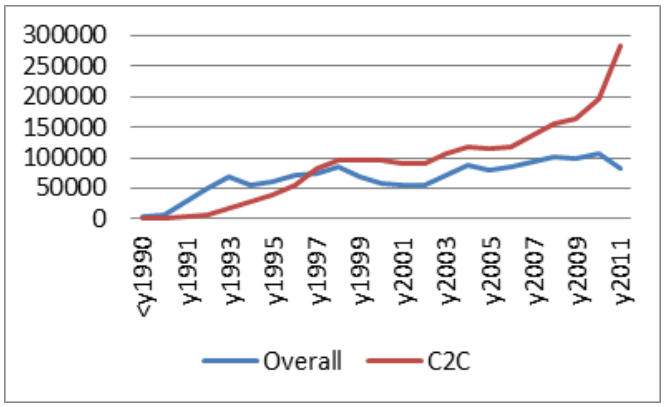


Fig. 2. New connections per year – Amout of connections, which were created in particular years. Overall connections mean number of connections in original SCN and C2C are connections from snapshots.

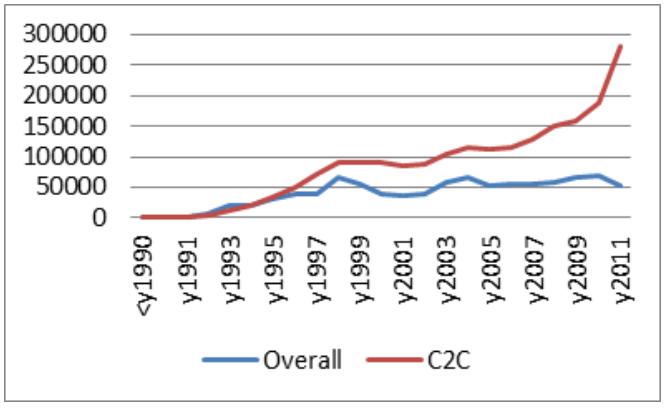


Fig. 3. Ending of connections per year – Amount of terminated connections during that year (connection was created before that year)

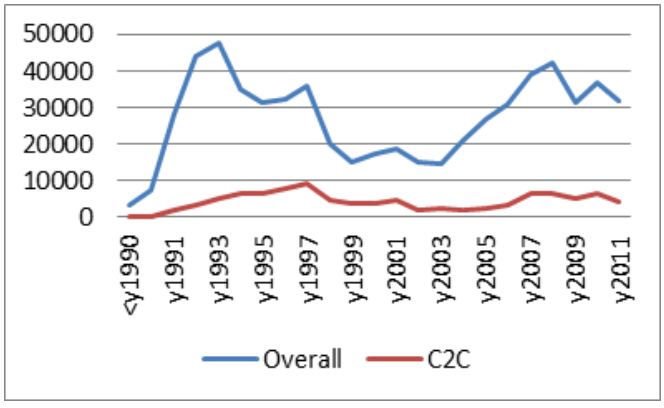


Fig. 4. Growth of connections per year – This graph describes "netto" increase of connections from year to year, which in fact reflects number of newly created connections (in that year) minus connections which ended.

At last **(5)** we create projected (one-mode) C-C Company Network originated from SCN. This projected network can be represented as simple undirected weighted graph and connections in the network represent collaboration between companies.

C. *Snapshots of Projected SCN*

One of the important parameters of temporal projection of SCN is the "now" date (i.e. date for which the snapshot should be generated). This date represents a time of projection, so the age of the connections is calculated to this time. Setting of this

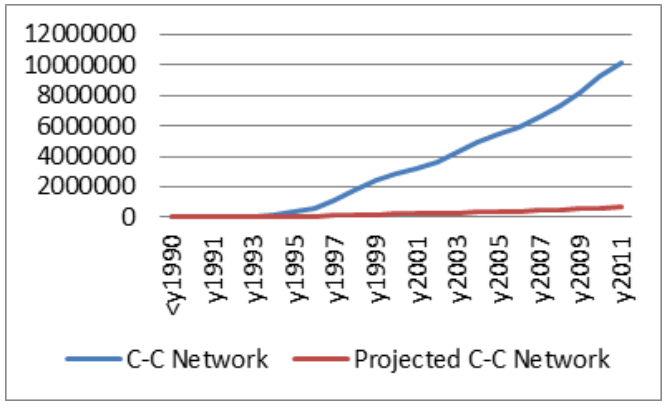


Fig. 5. Projected network - data downsizing. This graph shows decreased number of connections in projected network in time snapshots. C-C Network (blue curve) is networks in state on Fig.1. Step (3). Projected C-C Network (red curve) represents state on Fig.1. Step (5).

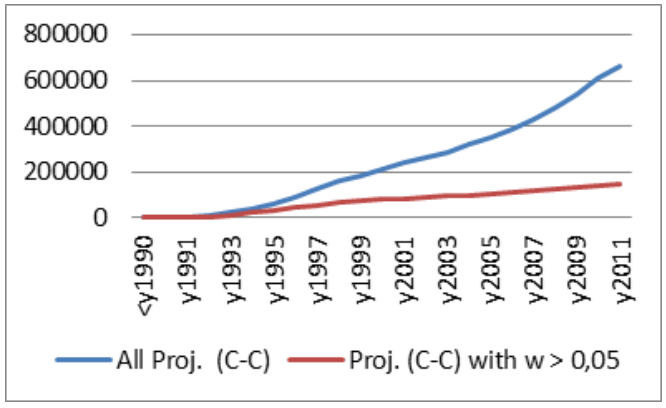


Fig. 6. Snapshot projections – example of state of alive connections. In this picture there are shown all projected connection in snapshots and connections with certain weight in snapshot to demonstrate the effect of aging of connections. As we see due the effect the number of "alive" connections does not grow as fast as the number of all connections.

parameter allows us to create infinite amount snapshots of SCN in time.

We have created 23 "year snapshots" of SCN as projected C-C network, where first snapshot is documenting state before year 1990 and next snapshots are generated for subsequent years, always to the date of 1st January in the year. The snapshot *y1990* reflects state of the network at the end of the year 1990 (on 31st of December 1990 inclusively).

Using these snapshots we calculated several statistics which describe some aspects of dynamics of SCN as you can see in the next section.

D. *Dynamics of Connections in Slovak Company Network*

Evolution of Slovak companies and business made breakthrough in 1990, where lot of dynamic changes started in business area. To this time relationship was really steady and no significant connections between companies were observable. From this time changes in ownership of companies and cooperation between companies increased in their number.

In figures (Fig. 2. – Fig. 6.) we can see statistics about connections in Slovak Company Network. In Fig. 2. and Fig. 3. we can see time dependency of creating and terminating connections of all types (any kind of connection of person actor P to company C or company C1 to company C2). Amount of these connections in SCN are represented by blue

curve, these are common new entries or changes in Business register of Slovak republic. From these connections projected connections are calculated and then they represent direct and indirect connections between companies.

Progress of creating projected connections between companies is depicted as red curve.

Conclusively these "Overall" connections (blue line) can be transformed into company to company connections by projections (red line). As you can see numbers of new "overall" connections per year are steady, but transformed (projected) C-C connections increase in amount rapidly as time goes. This is caused by the fact that the number of new companies is not increasing as fast as the number of new connections (connections are made between existing companies mainly by indirect connections). That means that with increasing of new connections the companies are more connected to each other (network is denser).

In Fig. 4 we can see comparison of overall connections growth and projected C-C connections growth. The situation is just opposite as it was in creation/termination of connections. The Dynamics of overall connections growth is fluctuant (you can see rapidly decreased growth in years 1999-2003) but C-C connections are quite solid. This shows, that even if a lot of connections is created or terminated in SCN, the projected C-C Network does not change rapidly. This is might be caused by the fact that lot of changes in ownership, management and other relationships of companies usually does not affect so much cooperation and collaboration between companies.

IV. Future Work

As our current work is very close to local structure analysis with temporal data, next future work will be examining of evolution in SCN. In the presented timeline (years 1990-2011) we can track common local structure and observe emerging patterns over time. For this purpose as part of current work we are developing method for heuristic decomposition of SCN and special block-modeling on local structure [9], [10].

V. CONCLUSION

In this paper we described temporal projection of specific social network (SCN) and pointed out some of its specifics. Mainly we contributed with analysis of its dynamic features and we created historical snapshots based on temporal data. This analysis can be very useful for researchers who are analyzing large dynamic networks, because of network data downsizing due to temporal projections.

Acknowledgment

This work was partially supported by the Slovak Research and Development Agency under the contract No. APVV-0208-10 (50%) and partially by the Scientific Grant Agency of the Ministry of Education, Science, Research and Sport of the Slovak Republic under the grant No. 1/1147/12 (50%).

References

[1] S. Wasserman, K. Faust, "Social Network Analysis", Cambridge University Press, ISBN 9178-0-521-38707-1, Cambridge, 1994

[2] P. Kostelník, P. Smatana, Itlis, "Transparent information in context," *Online at www.itlis.eu, 31-Oct-2011*

[3] M. Repka, J. Paralič, "Projections of Company Network with Temporal Data". In: CINTI 2012 : 13th IEEE International Symposium on Computational Intelligence and Informatics: In proceedings. Budapest, Hungary : Óbuda University, 2012 P. 19-24. - ISBN 978-1-4673-5204-8

[4] Ministry Of Justice Of The Slovak Republic, "Business Register On Internet," *Online at www.orsr.sk, 31-Oct-2011*

[5] M. Repka, "Analýza určitých typov sociálnych sietí. Písomná práca k dizertačnej skúške." Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky, Košice, Chapter 4, 2011.

[6] M. Repka, J. Paralič, "Projections of Company Network with Temporal Data", In Proceedings, 13th IEEE International Symposium on Computational Intelligence and Informatics, Budapest, Hungary, 2012

[7] M. Repka, J. Paralič, "Component Identification in Company Networks" In Proceedings, SCYR 2012 - 12th Scientific Conference of Young Researchers – FEI TU of Košice, Herľany, 2012

[8] M. Repka, „Local Structure Analysis in Company Network" In Proceedings SAMI 2012 Herľany, Slovak Republic. ISBN 978-1-4577-0195-5, 2012

[9] V. Bagatelj et al., "Generalized Blockmodeling with Pajek", Metodološki zvezki, Vol. 1, No. 2, 2004, p. 455-467

[10] V. Bagatelj, "Clustering and Blockmodeling", University of Ljubljana, Slovenia *Online at http://vlado.fmf.uni-lj.si/pub/networks/doc/KN/Course4.pdf*

[11] M.E.J. Newman, "Networks: An Introduction," Oxford University Press, Oxford, UK, 2010.

Event Announcer for Android Platform

[1]*Jakub ŠEVČÍK,* [2]*Ivan HALUPKA (2nd year)*
Supervisor: [3]*Jan KOLLÁR*

[1,2,3]Dept. of Computers and Informatics, FEI TU of Kosice, Slovak Republic

[1]sev.jakub@gmail.com, [2]ivan.halupka@tuke.sk, [3]jan.kollar@tuke.sk

***Abstract*—This paper presents the application for announcing acts for Android platform, which serves people in their day to day life. Paper demonstrates some examples of usage of our application, sums main difference between the day with usage of the application and a day without. We are also describing specific features of our application and their usage in every day life. In the end of the paper we describe our proposition of how to deal with upgrades, extensions as well as add-ons to all features of the application and also we are talking about new possibilities of application's usage.**

***Keywords*—android application development, scheduler, time app.**

I. Introduction

Nowadays, people cannot live without their smartphones and that also means without their applications. This is nothing new, applications in general are mostly helping us and making our lives easier. They are either providing us with some new information, saving us time and doing something for us (writing / sending messages, using voice dialing) and many more. This is something we would like to talk about, and more specifically our application called Scheduler. Scheduler is an application that is used to announce events that you predefined to be announced. It is similar to a basic calendar which can be found in any of the smartphones, but this application includes variety of additional features. In this age, men have to remember a lot of things which they should do during the day. Of course, it's not so easy to remember everything, this includes all of the age groups. For example, older people who have to take pills on daily basis or they need to go to the doctor often and can't remember any of this properly. Athletes can find usage for this application as well. It can be used to remind you training sessions, what to train and even where. It's very easy to use this application and that is its main advantage. Not many people like complicated stuff. Main usage for this application should be to remind you of something that you are doing on daily/weekly basis, not just birthdays remainder or so. For us, it was very important to make this application very easy to use, fully automated and independent of other applications. This means, if we have to do something more than once a day, application will automatically setup a proper settings and adapt to what you are doing with just a little help from you, which is adding a simple information about what are you doing at the moment.

We also implemented full support on how to setup everything manually if you wouldn't like how the application did it. This application will find usages mainly to people that are doing something often and on daily basis. We've targeted athletes and sick/older people. There is a very important feature and that's calendar. It calculates if people completed what they've setup. Another feature of this application is that you can back trace any day/week/month and see what you have completed and what not. All this data can be shown in percentage.

II. System structure

Structure of the application is created up to 3 main activities with personal layouts. You can move between those layouts by using swap across the screen either to the left or to the right side. There is also a dialogue window in the structure, which shows up on the home screen of your device with tree usable buttons (Delay, Info, OK). Services, which are run automatically when you boot your Android device, automatically log settings of the events and compare them with system time and date and based on this, you can see information in the dialogue window. Main important activities are:

A. Today

This activity represents main screen of the whole application after you run it. In the top of the screen you can see a header called "Today" together with the current date. You can find the number of total events for current day under the header "Today". On the bottom of the screen you can see all the events with the time when they should be completed along with the status as if it's completed or not. After you click on any of the events that are there, a new window will open up with additional information and that are description of the event, time of the event and status. There is also a button "edit", which let you add anything to the event so you will find it likeable or you can hit the button "delete" to delete the whole event. All events are being automatically saved in your local database where are automatically edited and then sent to a remote database server.

All the events that are shown under "Today" header are either colorized in white(this event was completed) or in red(this means when the remainder with the event came up you either hit "DELAY" button or you just simply ignored it). All this events are evaluated and later on they serve to show up statistics and some other information.

Fig. 1. Main screen of Scheduler's user interface

B. Calendar

This activity represents a calendar that shows the current month and the color resolution of the current day. When you want to see this screen, you need to use swap move to the left from the "Today" screen. When you will do the swap move to the right you will get back to the "Today" screen. After you hit any of the days in the calendar, you can open up the dialogue window that shows similar information to the "Today" screen. You can see the events that were not completed there as well. All those events are evaluated and after that you can see them on the main calendar screen in percentage. Based on this you can see how good you are at doing the stuff you predefine or have to do on daily basis. All this information is stored into the database. We have created a local and a remote database. Local database is used to store all the information for the current month. Remote database is used to store the older information. Reason for this is optimization between the device and the application. Device can be easily overloaded with data and because of that we have two databases so the device will run better.

C. Setting

This activity serves as header advice to edit the settings of the application. You can get to this screen by using swap move to the right from the "Today" screen. In the settings you can setup all the events that you need one by one. In other words, when you want to setup the events manually you have two options. If we don't want the event without a description, then we need to setup the brief description, status, also more in depth description (optional) time and then decide if the event should be done on daily basis or just on some of the days. If you will not choose whether the event should be done on daily basis or any other day, then it will be automatically setup to remind you this event on daily basis. All this information can be setup in the top screen of "Setting" screen. After you created an event you can find it in the calendar and if it's setup to remind you on daily basis, then it is also on "Today" screen. This is the first option how you can setup the settings. Second option is to create an empty event and then just edit it on your "Today" screen. You can also setup the application to run fully automatically. You need to add how many events you want and approximate time when you are usually going to sleep. After you fill out some additional information you will be able to add also how long are you usually sleeping as well as on which days you want to turn off the creation of automated events.

III. Automated settings

Application is automatically running with few options. First option is automated setup. This one works after you add the time when you are going to sleep (roughly) and after that this data is stored into the numerical condition and it will add the time of how much you sleep to it, which when you will not fill out will be setup automatically to eight hours. Based on this there are calculations regarding the time you have until you are going to sleep.

IV. Conclusion

As a closing statement we would like to mention various possibilities regarding the availability of add-ons and upgrades/updates in the future. One of the options is to create a system of profiles that works together with a web interface where the user could create their own account. Via this account they should be able to see their events also on the web interface that could be accessible from any of the internet browsers. For sportsmen, it would be very useful to add a feature of extensive events where they could log all their training plans and by that they could simply divide their specific training plan on a specific day. Next option would be to synchronize their events with a web interface. For example a doctor could set the events like taking pills, attending the health check in the web interface and that would sync up with the patient's application on the device that they are using so they would have all the events ready for use and they would not forget them so easily. This could also be very useful for trainers. They could create the whole training plans and nutrition plans and synchronize it to devices of their trainees. There are many possibilities for updates and upgrades in the future. Our goal in the development of this application was to help people and if this will be achieved we will be very satisfied.

References

[1] J.F. DiMarzio, *Android – A Programmer's guide*. New York: McGraw-Hill, 2008.

[2] R. Rogers, J. Lombardo, Z. Mednieks, B. Meike, *Android Application Development*. Sebastopol: O'Reilly Media, Inc., 2009.

Experimenting with Lazy Evaluation in Visualization

Csaba SZABÓ

Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

csaba.szabo@tuke.sk

***Abstract*—This paper deals with using the lazy evaluation property of a scripting language in visualization. The language chosen is Perl, as visualization core, OpenGL library is used. The power and weaknesses of using such a combination are shown on selected examples. The main goal is to point out the possibility to use a scripting language in the visualization part of a virtual reality system as well as it is known in e.g. the modeling or texture pre-processing phases of visualization preparation, respectively. The main feature of the selected scripting language used in the examples is lazy evaluation allowing on-demand visualization of the objects of the virtual reality scene.**

***Keywords*—Lazy evaluation, OpenGL, Perl, virtual reality, visualization.**

I. Motivation

Lazy evaluation is widely used in the implementation of functional programming languages [1], but the main idea of this method could find applications in many areas of computing. I.e. the language expression is being evaluated only when needed. The same idea is implemented in interpreted (e.g. scripting) languages, but in fact, the code of the script is also being pre-processed. Some interpreted programming languages such as Perl offer a more lazy property of source code processing that allows laziness in both syntax control and statement block evaluation (i.e. semantics determination).

Visualization [2] is the process of creating a visual output of a scene model representation. It could use 2-dimensional (2D) or 3-dimensional (3D) interfaces, respectively. The output could be a printed material or shown on an appropriate display. With the advent of Graphic Processing Units (GPU), realistic, real-time 3D rendering has become common. GPUs are designed to process large arrays of data, such as 3D vertices, textures, surface normals, and color spaces. OpenGL is an industry-standard, cross-platform language for rendering 3D images. Originally developed by Silicon Graphics Inc. (SGI), it is now in wide use for 3D CAD/GIS systems, game development, and computer graphics (CG) effects in film.

While OpenGL is in itself a portable language, it provides no interfaces to operating system (OS) display systems. As a result, Unix systems generally rely on an X11-based library called GLX; Windows relies on a WGL interface. Several libraries, such as GLUT, help to abstract these differences. However, as OpenGL added new extensions, OS vendors (Microsoft in particular) provided different methods for accessing the new APIs, making it difficult to write cross-platform GPGPU code.

There exist several approaches combining visualization and lazy evaluation, but these are based on lazy evaluation of expressions defining the scene itself or objects within the scene, respectively. In fact, geometric computations are lazy evaluated as in [3].

The main idea discussed in this paper points forward to the visualization of the scene that is mostly executed after all calculations related to the scene. The goal is to show the power offered by a scripting language with lazy evaluation of statement blocks of visualization code.

The organization of the paper is as follows. First, the lazy property of Perl is shown in Section II. Then, Section III continues with the discussion on perspective usage of the Perl-and-OpenGL language combination. Section IV contains basic examples of using the lazy property of Perl in visualization. Section V concludes and shows future directions of the research.

II. Perl's "Eval" Function

Perl allows explicit lazy evaluation of statements. Dynamic processing of user inputs is significantly different to this approach, because the `eval` function allows delay or denial of execution that is impossible when dynamically processing user inputs, which run in a so-called interactive scope. Lazy evaluation in Perl language is frequently used. There are two alternatives [4]:

eval EXPR In the first form, the return value of `EXPR` is parsed and executed as if it were a little Perl program. The value of the expression is first being parsed, and if there were no errors, executed in the lexical context of the current Perl program, so that any variable settings or subroutine and format definitions remain afterwards. The value is parsed every time the `eval` executes. This form is used to delay parsing and execution of `EXPR` until run time, that could be considered as the lazy property.

eval BLOCK In the second form, the code within the `BLOCK` is parsed only once – at the same time the code surrounding the eval itself was parsed – and executed within the context of the current Perl program. This form is typically used to trap exceptions more efficiently than the first one, while also providing the benefit of checking the code within `BLOCK` at compile time.

In both forms, the return value is the value of the last expression evaluated inside the mini-program. If there is a syntax error or run-time error, or a `die` statement is executed, `eval` returns `undef`, and `$@` is set to the error message. If there was no error, `$@` is guaranteed to be the empty string.

With an `eval`, there exist four cases to remember what is being looked at when:

```
1: eval $x;         # CASE 1
2: eval "$x";       # CASE 2
3: eval '$x';       # CASE 3
4: eval { $x };     # CASE 4
```

Cases 1 and 2 above behave identically: they run the code contained in the variable `$x`.

Cases 3 and 4 likewise behave in the same way: they run the code `'$x'`, which does nothing but return the value of `$x`. (Case 4 is preferred for visual reasons, but it also has the advantage of compiling at compile-time instead of at run-time).

In the field of VR systems, both alternatives of usage of the `eval` function could find their place – all four cases presented above. The second pair for exception-handling while the first one for lazy evaluation of the VR scene visualization code or its parts, respectively.

III. The OpenGL&Perl Perspective

Perl is very well suited for generating and rendering complex 3D objects and environments. It's built-in string and array handling functions make it well suited for manipulating large arrays of pixels and vertices – a common task for most 3D graphics applications. Several benchmarks have found Perl to be as fast, and often faster than other languages such as C and Python at many common types of complex rendering tasks.

A. Perl Bindings For OpenGL

Perl `OpenGL` (POGL) and `SDL::OpenGL` (Simple Direct Media Layer) are the two most developed Perl modules for rendering both 2D and 3D graphics, using the OpenGL libraries. The POGL module was found to perform significantly better than `SDL::OpenGL` in several benchmarking tests, when rendering OpenGL graphics[5], [6].

The POGL documentation is basically just a list of OpenGL functions that are supported, but it contains no definitions for these functions. This is not surprising, since all of the POGL functions work exactly like their C counterparts.

While Perl `OpenGL` outperforms `SDL::OpenGL`, it lacks some of the other capabilities that the SDL libraries possess. The two are often used together – POGL to deal with the OpenGL rendering, and SDL for things that OpenGL does not handle like setting up a display window, audio, and handling I/O events (mouse/keyboard/joystick/etc.).

B. Perl Advantages

Given that GPGPU performance will be a wash in most cases, the primary reason for using a compiled language is to obfuscate source for intellectual property (IP) reasons[7].

For server-side development, there is really no reason to use a compiled language for GPGPU operations, and several reasons to go with Perl are as follows[6]:

- Perl OpenGL code is more portable than C; therefore there are fewer lines of code
- Numerous imaging modules for loading GPGPU data arrays (textures)
- Portable, open source modules for system and auxiliary functions
- Perl (under mod-perl/ISAPI) is generally faster than Java
- It is easier to port Perl to/from C than Python or Ruby

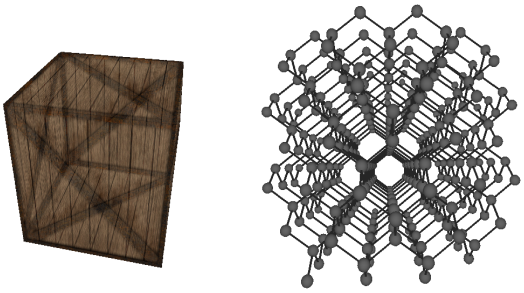


Fig. 1. A textured cube (left) and a look at the atomic structure of a diamond (right) displayed based on its PDB definition [9]

Desktop OpenGL/GPU developers may find it faster to prototype code in Perl (e.g., simpler string handling and garbage collection), and then port their code to C later (if necessary). Developers can code in one window and execute in another – with no IDE, no compiling – allowing innovators/researchers to do real-time experiments with new shader algorithms. Physicists can quickly develop new models; researchers and media developers can create new experimental effects and reduce their time to market.

Performance is not a reason to use C over Perl for OpenGL and GPGPU operations, and there are many cases where Perl is preferable to C (or Java, Python, or Ruby)[7].

IV. Calling OpenGL from "Eval"

This section presents a use case of using OpenGL with Perl without `eval` and various use cases for running OpenGL functions from an `eval` statement. The aim is to demonstrate the variety of possibilities those power and limitations will be discussed in the next section. The experiments were run on two test systems running FreeBSD OS with different Perl versions and hardware.

A. Experiment 1

The first experiment addresses the issue of static 3D object visualization. As the base, C++ code is used. The OpenGL related part of the code is translated into Perl and put into a POGL frame. Goals of the experiment are:

- to run all C++ examples in Perl,
- do not use any variant of `eval`.

Fig. 1 shows the outputs of two selected tasks:

1) The first one is the visualization of a textured cube using transparency and mouse control over object position, rotation, and distance.
2) The second task was to display the 3D structure of a larger molecule[1] that is stored in a file (file name is a command-line argument). The solution uses an object oriented front-end of POGL called `OpenGL::Simple`.

B. Experiment 2

Experiment two checks whether the running of dynamic scenes is also possible using Perl. Goals:

- to create an OpenGL application displaying two independently rotating objects both in C++ and Perl,

[1]The Protein Data Bank (PDB) is a repository for the 3D structural data of large biological molecules, such as proteins and nucleic acids. The file format initially used by the PDB was called the PDB file format. An XML version of this format, called PDBML, was described in 2005[8].

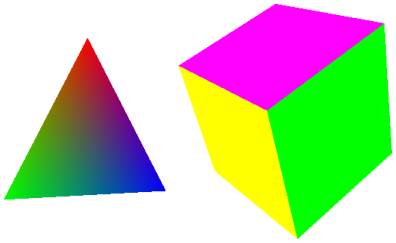
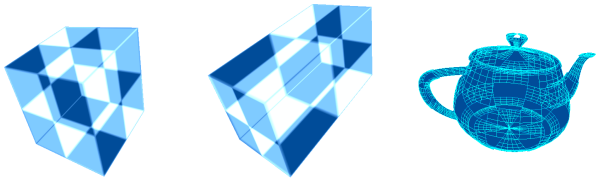


Fig. 2. A view at a dynamic scene



a) b) c)

Fig. 3. Example objects visualized using a simple `eval` statement

- do not use `eval` in the solution.

The result of the experiment is shown in Fig. 2. The cube and pyramid are rotating by two different speeds. Compared to previous solution, the implementation uses the basic POGL module.

First, the solution was developed using the `OpenGL::Simple::Viewer` module, but this module itself does not offer an interface to define `glutIdleFunc`. Fall-back to the basic POGL system for only dynamics showed conflicts in calls resulting into very slow motion of the animations and even slower user interaction. Therefore, the final solution was very identical to the C/C++ solution using only basic POGL.

C. Experiment 3

The third experiment is denoted to the usage of simple `eval` statements, i.e. ones creating a simple static scene, e.g. drawing a teapot. Static scene examples visualized using simple `eval` statements are shown in Fig. 3:

a) a textured semi-transparent cube,
b) the deformed version of the above cube object,
c) GLUT teapot object.

`OpenGL::Simple::Viewer` is used again here. The reason is the easier usage and the significantly shorter source code. The code frame is presented below:

```
1: ...
2: use OpenGL::Simple::Viewer;
3: ...
4: glutInit;
5: my $v = new OpenGL::Simple::Viewer(
6:  title => 'Window title',
7:  screenx => 512,
8:  screeny => 512,
9:  initialize_gl => sub {...}, #GL window
       initialization
10:  draw_geometry => sub {...}, #scene drawing
       code
11: )
12: ...
13: glutMainLoop;
```

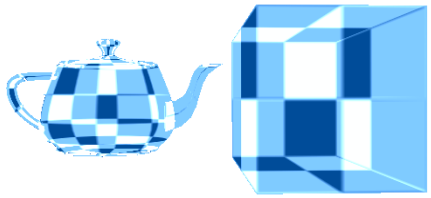


Fig. 4. Static scenes replacing the default one

The scene drawing code is being replaced in the experiment by using `eval` *case 2*. For demonstration the possibilities and similarities to the C/C++ solution, there follow two alternatives showing selection from existing drawing routines on one side (the C-like solution) and using simple code achieved as user input on the other:

```
1: print 'Enter number (1, 2, 3): ';
2: my $n = <STDIN>;          # Get input
3: chop $n;
4: unless ( $n > 3 ) {
5:  eval "\$v -> {'draw_geometry'} = \\\&gl$n;
       "; #select a routine
6:  die $@ if $@;
7: }
8:
9: sub gl3 {
10:  print 'Enter the command for draw_geometry
       : ';
11:  my $cmd = <STDIN>;       # Get input
       command for drawing, next
12:  chop $cmd;               # line replaces
       the drawing sub body.
13:  eval "\$v -> {'draw_geometry'} = sub {$cmd
       }";
14:  die $@ if $@;
15: }
```

D. Experiment 4

Experiment four enhances the functionality implemented in experiment three. Making it more efficient means allowing replacement of the basic static scene by another scenes. The basic static scene is the textured cube shown in Fig. 3 a). This one is being replaced by another code taken from the keyboard (simple input) or from a file. Because Perl also allows manipulation with variables such as introducing a new class attribute, the new scene could introduce dynamics into visualization. On Fig. 4, a new static scene from file is presented, while Fig. 5 shown a dynamic scene with a rotating pyramid.

To achieve the visualization of both scenes, new keys were introduced into the interface, e.g. 'F' to specify a new file-name with scene definition. To process the whole file, the following needs to be done:

```
1: if ( $c eq 'F' ) {
2:  print 'Enter file name: ';
3:  my $f = <STDIN>;          # Get input
4:  chop $f;
5:  open CODE, $f;
6:  undef $\;
7:  my @code = <CODE>;
8:  close CODE;
9:  eval "\$v -> {'draw_geometry'} = sub {
       @code };";
10:  die $@ if $@;
11: }
```

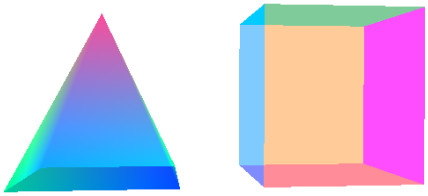


Fig. 5. Dynamic scene with two rotating objects replacing the default static scene

The already mentioned property of introducing new class members on-demand is used to create a dynamic scene presented on Fig 5. Sample code for introduction and further usage of the attribute `rotx` for rotation angle follows:

```
1: unless ( defined $v -> {'rotx'} )  {
2:  $v -> {'rotx'} = 0;
3: }
4: else {
5:  $v -> {'rotx'} ++;
6: }
```

V. Conclusion

The fact, that the experiments run on all systems with the same results, shows the power of Perl and OpenGL being OS-independent, but in the case of OpenGL only with the strong limitation of the OpenGL support of the graphics hardware and its driver.

In future, the plan is to run the experiments on Windows or other OS, respectively. The results might be used in the development of the Perl implementation of a VR system using various kinds of I/O [10].

Other possible application of the discovered behavior might be in the field of self-adaptive and/or self-healing systems that is also an area of our interest. In this case the laziness could be used for dynamic change of the application.

Acknowledgment

This work was supported by the Cultural and Educational Grant Agency of the Slovak Republic, Project No. 050TUKE-4/2013: "Integration of Software Quality Processes in Software Engineering Curricula for Informatics Master Study Programme at Technical Universities – Proposal of the Structure and Realization of Selected Software Engineering Courses."

References

[1] F. W. Burton and M. M. Huntbach, "Lazy evaluation of geometric objects," *IEEE COMP. GRAPHICS APPLIC.*, vol. 4, no. 1, pp. 28–33, 1984.

[2] Cs. Szabó, Š. Korečko, and B. Sobota, "Data processing for virtual reality," in *Advances in Robotics and Virtual Reality*, ser. Intelligent Systems Reference Library, T. Gulrez and A. E. Hassanien, Eds. Springer Berlin Heidelberg, 2011, no. 26, pp. 333–361.

[3] S. Pion and A. Fabri, "A generic lazy evaluation scheme for exact geometric computations," in *Library Centric Software Design (LCSD)*. Portland, Oregon: États-Unis, 2006. [Online]. Available: http://hal.inria.fr/inria-00344960/PDF/lazy-kernel.pdf

[4] *Perl Programming Documentation*. [Online]. Available: http://perldoc.perl.org/

[5] SuicideJunkie. Coding for OpenGL "live". [Online]. Available: http://www.perlmonks.org/?node_id=921909

[6] Graphcomp. POGL – a portable Perl binding for OpenGL. [Online]. Available: http://graphcomp.com/pogl.cgi?v=0111s3m1

[7] ——. POGL benchmarks – C vs Perl. [Online]. Available: http://graphcomp.com/pogl.cgi?v=0111s3B1

[8] J. Westbrook, N. Ito, H. Nakamura, K. Henrick, and H. M. Berman, "PDBML: the representation of archival macromolecular structure data in XML," *Bioinformatics*, vol. 21, no. 7, pp. 988–992, 2005.

[9] P. May. Diamond - molecule of the month. [Online]. Available: http://www.bris.ac.uk/Depts/Chemistry/MOTM/diamond/diamond.htm

[10] F. Hrozek, B. Sobota, Cs. Szabó, and Š. Korečko, "Augmented reality application in parallel computing system," in *7th International Workshop on Grid Computing for Complex Problems*, Bratislava, 2011, pp. 118–125.

Fast low-cost CNC position system

[1]*Jaroslav LÁMER (1st year),* [2]*Ivan KLIMEK (3rd year)*
Supervisor: [3]*František JAKAB*

[1]Dept. of Electrotechnics and Informatics, FEI TU of Košice, Slovak Republic

[1]jaroslav.lamer@tuke.sk, [2]ivan.klimek@tuke.sk, [3]frantisek.jakab@tuke.sk

***Abstract*—This paper describe theoretical and practical optimization and build of electro-circuits for low-cost CNC devices with emphasis on low price. Device is based on accessible parts like ATMega8 or L298. It accepts post-processed G-Code communication protocol. Communication with controlling computer is based on USB interface and is not native. Used is FT232 USB to RS232 converter circuit. On computer side of communication is running custom or third party software for communication via serial interface. Optimizations are made on both of software and hardware parts of system. The small size of the device microprocessor program (8kB) was achieved using algorithms for working with integer data-types and with abstraction of floating point data-types operation into a controlling computer.**

***Keywords*— Optimization, CNC devices, Atmega8, L298, low-cost, CAD, CAM, post-processor. 8kB G-Code parser**

I. Introduction

The area of CNC devices contains of wide spectrum of tools and methods for computer mechanism driving. The main problem in amateur, homemade and small-scale factoring is its high price, unavailability of some electro-technical parts, etc. Therefore more users often use his own experiences and skills and made CNC herself. Solution for mechanical construction is for each user individual, but the electronic and electro-driving circuits are the same. This fact makes electronic area of CNC device construction the best candidate for physical, or price optimization. Physical and price optimization are self-depended.

II. Function analysis of CNC devices

First prerequisite for driving any CNC device is the idea or imagination of an object, that will be constructed. For model designing are used CAD tools or vector editors. For example Autocad, SolidEdge, Inkscape, etc. CAD tool output is computer model of object, characteristic by format of used tool. The CAD model is just mathematical object. It needs to interpret as set of moving or position instructions. This instructions executed in CNC device finally leads to physical movement. It is possible to use many types of technologies for making the physical objects. For example milling, drilling, laser cutting, etc. Object movement interpretation is result of CAM software tool. This tool makes transformation from computer model to physical moving instructions known as G-code in automatic mode, or semi-automatic mode with manual final code correction. It is possible to execute G-code on CNC device directly. G-code is the communication protocol between computer and CNC device. [1]

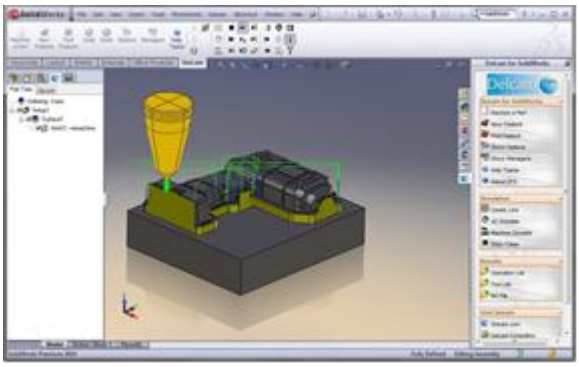


Fig. 1 Example of CAM software [2]

It is based on codes. Each base code begins with G (they are many types of code like G,M, etc.). Codes are divided into lines (dividing by end of line character). Each code represents one operation. There are many types of codes, for example here is the template and the most used codes.[1]

CODE [**PARAM1**value][**PARAM2**value][...]

G0 Xddd.ddd **Y**ddd.ddd **Z**ddd.dddd– fast movement to given position
G01 Xddd.ddd **Y**ddd.ddd **Z**ddd.dddd– linear interpolation
G02 Xddd.ddd **Y**ddd.ddd **Z**ddd.dddd **J**ddd.ddd **K**ddd.ddd– clockwise circle interpolation
G20 – switch to inch unit system
G21 – switch to mm unit system

Basic codes start with G character with its number, parameters and parameters numerical values. Next are coordinates given in mm or inches (G21 and G20 switch between unit systems). Coordinates are given in float data-type. Communication between PC and CNC may be based on RS232 serial interface (it's possible to use the USB interface with RS232 emulator and converter) or LPT parallel interface.[4] Computer sending codes to CNC device in that order, in what they are generated by CAM tool. Codes are therefore a kind of vector description of geometrical shapes. All codes in order represent physical procedure to make the CAD-designed model. The central microprocessor in CNC device receives the code, calculate dimensions and distances

and execute the rasterization algorithm.[5]

In the low-cost devices are often used microprocessors from Atmel.[6] For example ATMega microprocessor set. G-code rasterization is computed in the field of float values. This implies demand to big program flash memory in central microprocessor. Existing solutions work with Atmega168 chips or better, they have more than 16kB of internal flash memory. In retail is it hard to find and buy. The most accessible chips are ATMega or ATTiny with programmable flash memory up to 8kB of size. Next image shows the standard data-flow process from CAD design to physical motor driver.

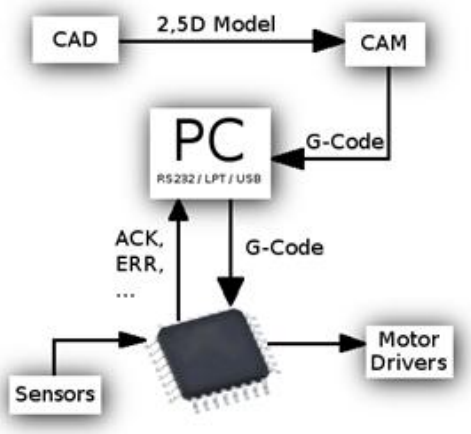


Fig. 2 Data-flow process from CAD design to physical movements of CNC axis

In CNC devices are frequently used the stepping DC or AC motors.[6] These motors have many I/O pins. Combination of polarities on this pins change rotor rotating position – also known as step (stepping). Physical angle of step is motor-type depended and it's defined by manufacturer. It's necessarily to compute the relation - number of steps / shift the axis in distance unit. This constant is computed for each motor by architect only at device design time. Constants may be all the same for all motors (all motors have the same number of steps for one turn of rotor and the same gear to axis), or may be all different. Constants are depended on motor operation mode too.[5]

It's possible to operate in half-step mode, or in full-step mode. In full step is possible to turn quickly, but the rotor is susceptible to slip (executed number of steps is not physically executed on motor). Half step mode is slower, but more precise and safe. When device executing G-code, dimensions in g-code (mm, inch) must be transformed to number of motor steps. Number of steps represent G-code dimension, transformed to axis movement.

Motor driving is performed by integrated power driving circuits (H-bridges), they are compatible with TTL logic system, or by special PCB mounted circuits, based on MOSFET transistors (manually build H-bridges).

One example of H-Bridge is integrated power circuit L298, or L293D. L298 is designed for driving with currents up to 2A, L293D up to 1A. These circuits drive motor voltages in range of tens volts (depended of type and method of use). It's possible to connect driving circuits to microprocessor directly, or using adaptation circuit (it reduce count of communication wires between microprocessor and driver circuit). Example of adaptation circuit primary designed for communication with L298 is L297.[7] [10]

This driving system is applied on all axes. For security reasons and for calibration needs, margin sensors are used. It's possible to use many types of sensors working in TTL level system from physical micro-switch sensors to optical sensors. One kind of sensor is an optocoupler. Switching between states is made by interrupt of infra beam between infra led diode and infra transistor. The main advantage of this sensor is its resistance to physical depreciation. One of optocoupler disadvantage is need to power supply.

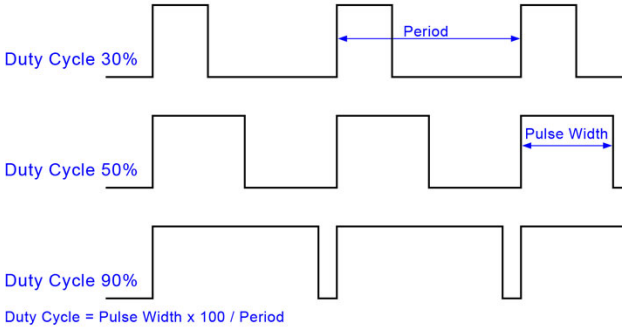


Fig. 3 PWM principe diagrams [3]

Power regulation of milling or drilling tool is based on PWM (pulse width modulation) modulated regulators or relay based switching parts and circuits. Relay based drivers have one big disadvantage – just two power state. PWM based regulators can modify driving power in wide spectrum. PWM based drivers disadvantage is the power supply need (in hi-power application cooler need). [8]

A. *Analysis and solution of possible improvements*

Theoretical analysis show more ideal candidates for price or availability optimization. In refer to main targets of this paper – availability of CNC devices for spread spectrum of public amateur users, possibilities to use the cheap and available parts are discovered. The main problem is replacement of driving microprocessor with sufficient programmable flash memory and simultaneously with sufficient count of I/O communication pins.

The main available microprocessor from Atmel is ATMega8 with 8kB of programmable memory and three (8bit) communication gates. [9]

It is possible to reduce count of used part for axis driving. This reduce price of driving circuits about 50%. In new computers is hard to find any RS232 or LPT communication interface. They were replaced by universal serial interface (USB). Communication line need to be converted to USB interface natively, or using a converter circuit.

III. Reduction of Processor Flash Memory Requirement

In case of program memory reduction and wide rasterization algorithms, working with integer numbers is needed. It is possible at the expense of precision, or using the step position system. Second option is friendlier, because the precision stay at the same level and because converting to step system is executed always after rasterization in any CNC

machine. This means, that conversion from metric or imperial units to stepping dimension system is moved from CNC device to computer. So three issues were removed or moved to computer memory:

1. Conversion from text to float values – conversion is from text to integer value
2. Computing in float values - computing executing on integer values
3. Conversion between dimension systems – moved from CNC to PC

The solution to do computer side dimension conversion is to use G-code post-processor – a program, which is executed after generating a G-code with CAM tool. It converts unit system to step system. Is possible to create new software for these purposes, or use existing solutions for post-processing.

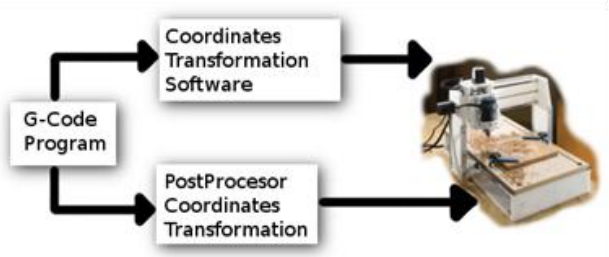


Fig. 4 G-code transformation options

In microprocessor rasterization algorithms, those work with integer dimensions, are needed. Ideal algorithm for three dimensional linear interpolation is Bresenham algorithm. For circle interpolation is ideal Bresenham algorithm too, but need to be modified to draw objects continuously.

In hardware part of system is possible to optimize count of used chips – concrete a chip for communication between microprocessor and driving circuit (L297).[10] With this reduction, number of communication pin increase from four, to six. Analysis of communication pins logical values is shown on next image. It explains that two pins are logically depended on other four. Therefore these two pins can be directly computed.

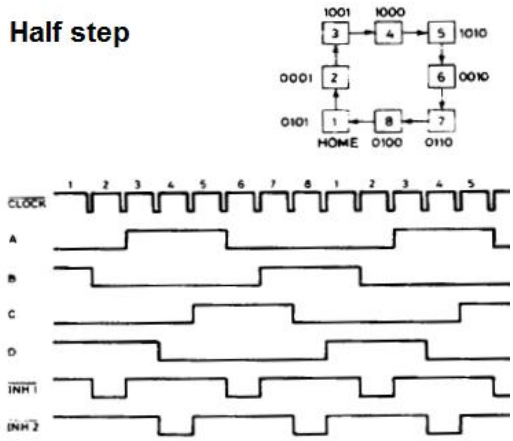


Fig. 5 Half-stepping mode diagram [10]

Image shows, that exists dependency between pins A,B,!INH1 (dependency is equal to C,D,!INH2). This dependency can be explained as:

!INH1 = A | B

!INH2 = C | D

Connection and communication between CNC and computer can be done using USB interface.

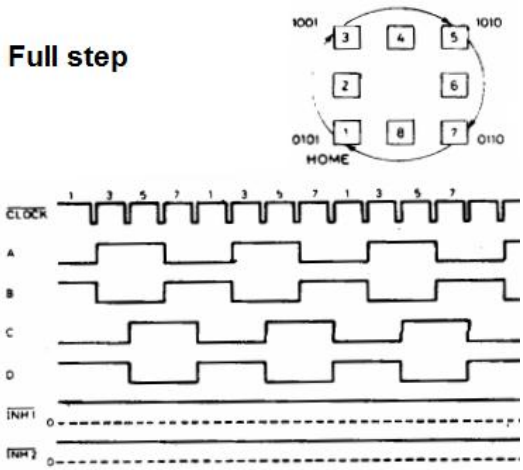


Fig. 6 Full-stepping mode diagram [10]

Microprocessors ATMega8 haven't this interface. They have only RS232 (USART), TWI and SPI interface.[9] Problem is solvable using converter from RS232 to USB. For example can be used FT232RL.

Fig. 7 PC to CNC connection

To controlling computer is this converter connected directly through USB port and can be found in device manager as COM serial communication port. This is great feature for compatibility with existing software products, those using this interface for communication through RS232 serial communication interface available on all older computers. Converter driver for PC is distributed by its manufacturer and its available for many platforms! In case of writing own software for communication between PC and CNC device, standard serial port libraries available in VisualStudio or QT can be used.

IV. Conversion between distance systems and rasterization

Post-processor and computer-side program is made as one software written in C++ programming language (used multi-platform C++ QT programming environment). Software reads G-code file with codes and transform it from used distance system to step distance system. These codes are then sequentially posted to CNC device. Software expects submitting execution of posted codes from CNC device side. Progress of send and code execution is logged to separate file. This feature simplifies reverse control of executed code in case of communication or debugging error.

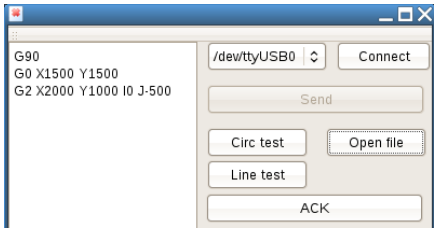


Fig. 8 Coordinate system transformation software

In microprocessor were implemented algorithms for linear interpolation and circle interpolation of Bresenham type. The final optimization of CNC device and communication model is showed on next image. In red are differences.

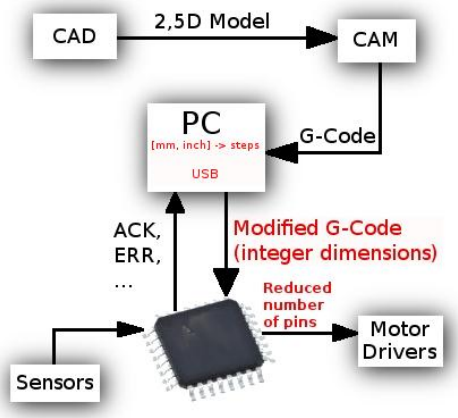


Fig. 9 Final optimalisation of model described on figure 2

V. Electronic Solution

The core of CNC device is placed by Atmel ATMega8 microprocessor. It doesn't have required count of I/O pins. Therefore hardware optimization is made. Four pins are maximum available for driving each device axis on used microprocessor. Driving circuit needs six communication pins. This is solvable by directly hardware computing of any two pins. As is explained in optimization chapter, the logical dependency of computed pins is equivalent to logical function TTL OR. The easiest way to compute this function is to use a DL (diode logic) logic and its equivalent to TTL OR = DL OR. [11] This logic computing system use only diodes (cheap parts).[11] Physical connection of DL OR shows next image. Output 1 = input 3 | input 4.

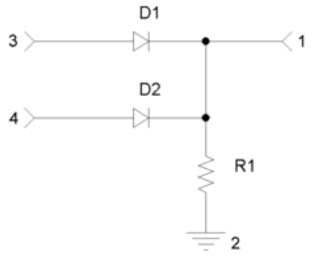


Fig. 10 TTL OR function in diode logic [11]

RS232 to USB converter output have power voltage pins, obtained from USB interface, so this is usable in power supplying of whole device including sensors. The only needed external supply voltage is supply for motor drivers – motor driving voltage.

VI. Conclusion

In this paper was successfully analyzed and documented optimized solution of CNC device electronic and software driving, using accessible technologies with emphasis to final device price reduction. The final prototype is made and fully functional. Comparing to other solutions is cheaper, simpler and meets all requirements given to this kind of device. But for lack of I/O pins, small count of external sensors is possible to connect to device. For example, in compare to other CNC devices, motor temperature sensors or overloading of milling motor sensors is missing. It's possible to improve whole of the device. Example of great improvement is motor slip detection using optocoupler. This will solve the main stepper motor disadvantage and increase working precision. System will be equivalent to sensors in computer mouses, based on ball gear. On axis will be added performed circle. Number of perforation will be a half of steps number of motor, used on this axis (in case of full-step driving mode). In case of half step driving mode will be count of perforation equivalent to count of motor steps used on that axis. Step (half-step) will be detected as logical level switch on optocoupler sensor. Next image shows this improvement. Red color represents optocoupler sensor, blue the axis rod. Black represents the performed circle.

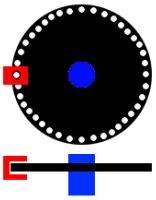


Fig. 11 Possible improvement - motor slip detection

Acknowledgment

This work is the result of the Project implementation: Competency Centre for Knowledge technologies applied in Innovation of Production Systems in Industry and Services, ITMS: 26220220155, supported by the Research & Development Operational Programme funded by the ERDF.

References

[1] ALTINTAS, Yusuf. Manufacturing automation: metal cutting mechanics, machine tool vibrations, and CNC design. Cambridge university press, 2012.
[2] http://down.cd/images/apps/SolidCAM-2011-with-SP2-HF1-for-SolidWorks-2007-2012-6581.jpg
[3] http://d32zx1or0t1x0y.cloudfront.net/2011/06/atmega168a_pwm_02_lrg.jpg
[4] DNC Software Ltd, Quick and easy to use CNC communication software, 2013,[Online; Accessed 11-March-2013], [Online], Available: http://www.dnc4u.com/
[5] Art Fenerty and John Prentice, Používání Mach3Mill, 2003, ,[Online; Accessed 11-March-2013], [Online], Available: http://www.cnc1.eu/files/images/stories/mach3/Mach3Mill_CZ.pdf
[6] http://www.c-n-c.cz/index.php
[7] STMicroelectronics, L298, 2000, [Online; Accessed 11-March-2013], [Online], Available: http://www.sos.sk/a_info/resource/c/stm/l298.pdf
[8] BARR, Michael. Pulse width modulation. Embedded Systems Programming, 2001, 14.10: 103-104.
[9] http://www.datasheetcatalog.org/datasheet/atmel/2486S.pdf
[10] STMicroelectronics, L297, 2000, [Online; Accessed 11-March-2013], [Online], Available: http://www.datasheetcatalog.org/datasheet2/c/0g0gh362kxwx30o7xi4x2i4jzffy.pdf
[11] http://hyperphysics.phy-astr.gsu.edu/hbase/electronic/diodgate.html

Feature extraction methods for the robust acoustic event detection system

Eva KIKTOVÁ (4^{th} year)
Supervisor: Anton ČIŽMÁR

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

eva.kiktova@tuke.sk

***Abstract*—With the increasing use of audio sensors in surveillance or monitoring applications, the detection of acoustic event performed in a real condition has emerged as a very important research problem. In this paper we investigate the impact of changing SNR condition from -3dB to 20dB. This paper brings also the comparison of different feature extraction algorithms which are inspired by algorithms used in speech processing applications. Three types of features were used for description of background sounds and acoustic events such as breaking glass and shots. Hidden Markov model (HMM) based learning technique performs the classification of mentioned sound categories. Coefficients of mel-spectrum (MELSPEC) or logarithm of mel-spectrum (FBANK) with first, second time derivations and cepstral mean normalization (_DAZ) outperformed the recognition performance of conventional Mel-frequency cepstral coefficients (MFCC_DAZ).**

***Keywords*—Acoustic event detection, HMM, SNR.**

I. Introduction

The acoustic event detection is currently very attractive research domain. The term of acoustic event denotes the specific sound category, which is relevant for the particular task. It usually has a rare occurrence and it is hard to predict when and whether it occurs. This paper is focused on two sound classes, i.e. gunshot and breaking glass. These sounds are relevant for security applications and they belong to the foreground sounds, which appointed to an abnormal situation. In normal conditions, the foreground sounds do not occur, but when they occur, it probably determines some criminal activity.

The intelligent security (or surveillance) system works according to the given instructions and it is not influenced or limited by various factors, not as a human. It should work autonomously and generate alert only when some dangerous situation is detected. It should also provide a constantly support for the police operator.

The detection of acoustic events is a partial task in the complex security system [1].

The detection of acoustic event in real condition is a challenging task [2], [3], [4]. An unstable background sound, different weather conditions, changing values of SNR and many similar noisy non-event sounds (like trucks, trams, etc.) limit the performance of each security system. In this paper we investigated the impact of very important limiting factor - the noise level measured as Signal-to-Noise-Ratio (SNR).

Different types of audio features (MELSPEC, FBANK and MFCC) [5] and HMM prototypes (different number of states) were tested. Finally, we identified the suitable parametric representation and HMM models for the acoustic event detection in noisy environments. Also the appropriate procedures for acoustic event feature extraction task were determined.

The rest of the paper has the following structure: Section 2. gives information about applied feature extraction methods and Section 3. gives information about the used part of sound database. Section 4. describes performed experiments and Section 5. summarizes obtained results, then follows the conclusion is Section 6.

II. Feature Extraction

The feature extraction is a crucial aspect for each detection system, because the recognition performance depends on the quality of extracted feature vectors. In this work Mel-frequency cepstral coefficients (MFCC), logarithmic mel-filter bank coefficients (FBANK) and mel-filter bank coefficients (MELSPEC) were used [12], [13]. Their computation process is depicted in the Fig. 1.

Acoustic signal is filtered using preemphasis filter then the 25ms Hamming window method was applied on the frames. Then, they are transformed to the frequency domain via the discrete Fast Fourier Transform (FFT), and the magnitude spectrum is passed through a bank of triangular shaped filters. After this step we have MELSPEC features. The energy output from each filter is then log-compressed and represent FBANK features. Finally, MFCC coefficients were obtained after the transformation to the cepstral domain by the Discrete Cosine Transform (DCT). As it was described earlier, the MELSPEC coefficients represent linear mel-filter bank channel outputs and FBANK coefficients are logarithmic mel-filter bank channel outputs. Mentioned features are computed during the extraction of widely used MFCC coefficients.

III. Acoustic Event Database

The extended acoustic events database JDAE TUKE [10] used in this work involved the gun shot recordings with 463 realisations of shots from commonly used weapons, breaking glass recordings with 150 realisations of broken glass and background recording (traffic sounds) with the duration of 53 minutes.

SNR influence was investigated on the recordings with different levels of SNR (Signal Noise Ratio). Nine new recordings (approximately 53 min) with different SNR i.e. -3dB,

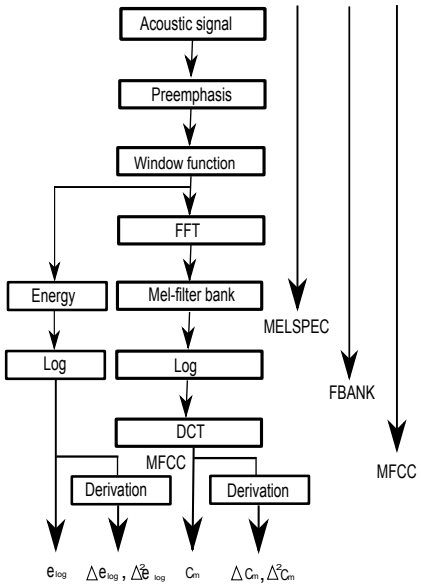


Fig. 1. Principial block scheme of MELPSEC, FBANK and MFCC coefficients.

0dB, 3dB, 6dB, 8dB, 11dB, 14dB, 17dB, 20dB were created and used in the training process.

In the testing phase 33 min recordings with different SNR were recognised. They contained two series of continuous gun shots and two series of continuous breaking glass events. Shots and breaking glass sounds formed one class i.e. the acoustic events class, so they were evaluated together.

All recordings had wav format with 48 kHz sampling frequency and the resolution of 16 bits per sample. Recordings were manually labeled using Transcriber software [11].

IV. Description of Experiments

Many internal experiments with speech based features such as Mel-Frequency Cepstral Coefficients (MFCC), Perceptual Linear Prediction (PLP), Linear Prediction Coefficients (LPC) and Linear Prediction Cepstral Coefficients (LPCC) were done. Relatively good results were obtained by MFCC coefficients, which achieved better recognition performance than the rest of the mentioned parametrization approaches.

Features were extracted without a preemphasis filtration. 25 ms of Hamming window and 10 ms of frame step were used. 22 Mel-filters were applied and the final number of coefficients was set to 12 (or 13 when using also energy or zero cepstral coefficient). Each acoustic HMM model was evaluated by the classifier based on the Viterbi decoder. Experiments presented in this paper were performed in HTK (Hidden markov model ToolKit) environment [5].

For this reason we decided to investigate MFCC approach more precisely. We evaluated different settings of MFCC with, zero cepstral coefficient (0), energy (E), delta (D), acceleration (A) coefficients and with or without cepstral mean normalisation (Z).

Generally the presence of energy E or 0th coefficient brought significant improvement in the relatively quiet environment without any louder sounds and with stable value of SNR. But, it is very difficult to restrict louder sounds and to have appropriate level of SNR in a real environment.

First and second time derivations of the baseline coefficients expanded feature vectors from 12 to 36 coefficients (or 13 to 39). The derivated coefficients had positive impact for each tested scenario. Generally speaking, also HMMs with higher number of PDFs achieved significant improvements.

Acoustic models trained with MFCC_EDA or MFCC_0DA features have in a testing process low detection performance because tested input sound had different SNR ratio (between -3dB and 20dB). Created HMM models did not correctly recognized input sounds. MFCC_EDA and MFCC_0DA models were successfully applied only for particular SNR value of tested audio signals.

Cepstral mean normalisation (CMN) [5], [9] is very effective for the recognition performed in a noisy environment and it is de facto standard operation for most large vocabulary speech recognition systems. The CMN algorithm computes a long-term mean value of the feature vectors and subtracts the mean value from the all feature vectors. CMN reduces the variability of the data and allows simple but effective feature normalization. CMN in HTK is realized by the optional value "Z" [5].

On the basis of promising results with MFCC_DAZ, we decide to investigate also features, which were obtained before final feature set. Therefore MELSPEC_DAZ and FBANK_DAZ were extracted too. The impact of the logarithmic operation and decorrelation by DCT was also evaluated.

In these experiments ergodic HMMs from one to four states and from 1 to 1024 Probability Density Functions (PDFs) were trained and evaluated in offline tests.

V. Results of Experiments

Three types of features MELSPEC_DAZ, FBANK_DAZ and MFCC_DAZ were used for the evaluation of the proposed robust system. The used testing set consisted from 9 recordings, more details are available in the Section III. Obtained results of experiments are depicted in the Fig.2, where the SNR results were averaged for each model type using Accuracy (ACC%) measurement technique [5], [6].

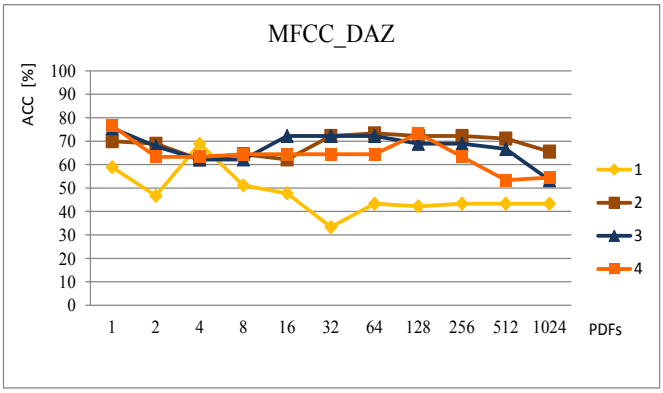
The approach based on the MFCC_DAZ parametrization achieved results higher than 75% only in the two cases, i.e. ACC = 76,67% was achieved by four states HMM with 1 PDF (HMM_4_1) and ACC = 75,56% with HMM_3_1.

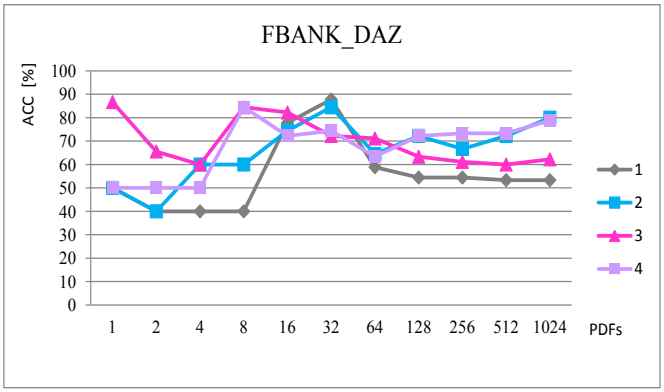
FBANK_DAZ features obtained promising results. The accuracy higher than 75% were achieved nine times. The highest value of ACC=87,78% was reached by HMM_1_32, second best results ACC=86,67% belonged to HMM_3_1. Like a previous case, MELSPEC_DAZ features were evaluated by the same way. Four times ACC higher than 75% were yielded. The best MELSPEC models HMM_1_64 achieved 80% ACC.

Generally suitable results were yielded by FBANK features, otherwise MFCC features seem to be the less appropriate. As it was mentioned, the results from Fig. 2 were averaged.

By averaging of results depicted in the Fig. 2 we lost information about the recognition of praticular SNR recordings, but we obtained information about the robustness of given approaches (MFCC_DAZ, FBANK_DAZ, MELSPEC_DAZ). Some models achived 100% ACC for high SNR, but for SNR=-3dB or 0dB the obtained ACC was low. Therefore we detailed analyzed all achieved results and identified three best models.

In the case of MELSPEC_DAZ models that reached mediumly ACC=70%, relatively low ACC for SNR -3dB and 0dB were achieved, but for high SNR the perfect recognitions were





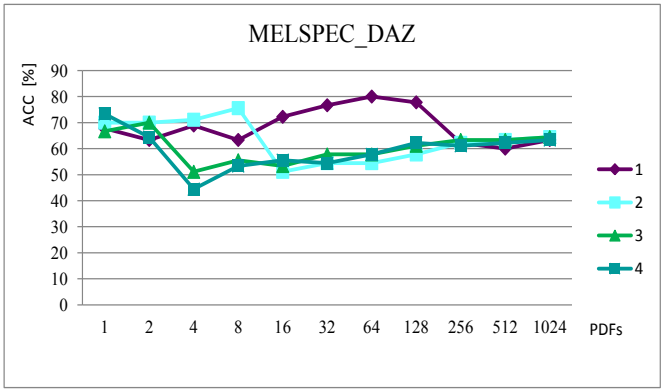


Fig. 2. The recognition results of average ACC [%] for each tested HMM model. (HMM_state_PDFs).

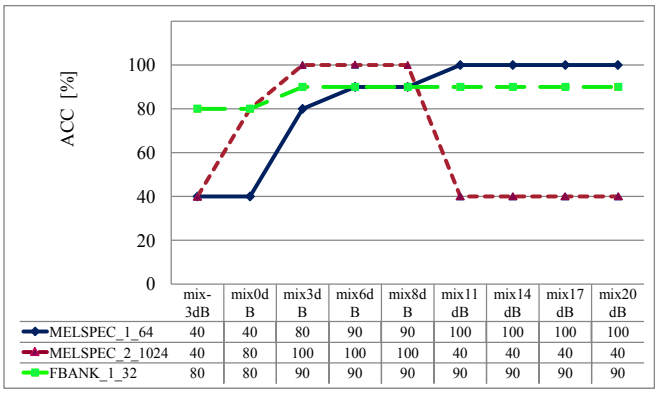
yielded several times. The same trend was occurred very often, therefore we analysed also other approaches (MFCC_DAZ and FBANK_DAZ) for finding the most suitable models regarding to the particular SNR.

The most balanced results were achieved by the FBANK_DAZ where the recognition results for low and high SNR were in the range from 80% to 90% of ACC. These results are depicted in the Fig. 3.

VI. Conclusion

This paper evaluated the acoustic event detection in the urban environment in consideration of the very important limiting factor - the noise level measured by SNR.

We analysed the detection performance for different SNR conditions with using MFCC, FBANK and MELSPEC features. Enhancements such as delta, acceleration coefficients and cepstral mean normalisation were applied as robust features for acoustic events recognition.



	mix-3dB	mix0dB	mix3dB	mix6dB	mix8dB	mix11dB	mix14dB	mix17dB	mix20dB
MELSPEC_1_64	40	40	80	90	90	100	100	100	100
MELSPEC_2_1024	40	80	100	100	100	40	40	40	40
FBANK_1_32	80	80	90	90	90	90	90	90	90

Fig. 3. The best achieved recognition results for different SNR ratio.

Performed experiments showed that MELSPEC_DAZ and FBANK_DAZ features are able to distinguish the presence of acoustic events in different SNR conditions more accurately than widely used MFCC_DAZ. Promising results especially for MELSPEC_DAZ were achieved by combination of two models (1_64 and 2_1024). FBANK_DAZ provided the most balanced recognition for all SNR. Delta, acceleration coefficients and cepstral mean normalisation contributed apparently to the robustness of created acoustic models.

Acknowledgments

This work has been performed partially in the framework of the EU ICT Project INDECT (FP7 - 218086) and under research project ITMS-26220220155 supported by the Research & Development Operational Programme funded by the ERDF.

References

[1] INDECT project homepage, http://www.indect-project.eu/
[2] Clavel,C., Ehrette, T., Richard, G.: Events Detection for an Audio-Based Surveillance System. In:IEEE International Conference on Multimedia and Expo 2005, pp. 1306–1309 (2005)
[3] Atrey, P. K., Maddage, N. C., Kankanhalli, M. S.: Audio Based Event Detection for Multimedia Surveillance. In: IEEE International Conference on Acoustics, Speech and Signal Processing 2006, vol. 5, pp. 813–816 (2006)
[4] Pleva, M., Lojka, M., Juhar, J., Vozarikova, E.: Evaluating the Modified Viterbi Decoder for Long-Term Audio Events Monitoring Task. In: 54th International Symposium Croatian Society Electronics in Marine - Elmar 2012, pp. 179–182 (2012)
[5] Young, S., et. al.: The HTK Book. Cambridge University, pp. 368 (2006)
[6] Vozarikova, E., Lojka, M., Juhar, J., Cizmar, A.: Performance of Basic Spectral Descriptors and MRMR Algorithm to the Detection of Acoustic Events. In: Communications in Computer and Information Science : Multimedia Communications, Services and Security. no. 287, pp. 350–359 (2012)
[7] Vozarikova, E., Juhar, J., Cizmar, A.: Acoustic Events Detection Using MFCC and MPEG-7 Descriptors. In: Communications in Computer and Information Science : Multimedia Communications, Services and Security, vol. 149, pp. 191-197 (2011).
[8] Kim, H.G., Moreau, N., Sikora, T.: MPEG-7 audio and beyond: Audio content indexing and retrieval. Wiley, pp. 304 (2005)
[9] Aik Ming Toh, Togneri, R., Nordhoolm, S.: Investigation of Robust Features for Speech Recognition in Hostile Environments. In: Asia-Pacific Conference on Communications 2005, pp. 956–960 (2005)
[10] Pleva, M., Vozarikova, E., Dobos, L., Cizmar, A.: The joint database of audio events and backgrounds for monitoring of urban areas. In: Journal of Electrical and Electronics Engineering, vol. 4, no. 1, pp. 185–188 (2011)
[11] Transcriber, http://trans.sourceforge.net/
[12] Psutka, J., Müller, L., Psutka, J.V.: Comparison of MFCC and PLP parametrizations in the speaker independent continuous speech recognition task. In: Eurospeech 2001, pp. 1813–1816 (2001)
[13] Wong, E., Sridharan, S.: Comparison of linear prediction cepstrum coefficients and mel-frequency cepstrum coefficients for language identification. In: International Symposium on Intelligent Multimedia, Video and Speech Processing 2001, pp. 95–98 (2001)

From User Stories to Predicate Logic in Requirement Representation

[1]*Veronika SZABÓOVÁ (1st year),* [2]*Emília DEMETEROVÁ (1st year)*
Supervisor: [3]*Zdeněk HAVLICE,* [4]*Valerie NOVITZKÁ*

[1,2,3,4]Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

[1]veronika.szaboova@tuke.sk, [2]emilia.demeterova@tuke.sk, [3]zdenek.havlice@tuke.sk, [4]valerie.novitzka@tuke.sk

***Abstract*—Adequate specification of the requirements is a very important knowledge for building successful software projects. This paper analyses three aspects of requirements representation: user story, model of requirements, and predicate logic; and proposes new relations between them. We also present our proposed method of formal representation of knowledge from user stories through requirements model to predicate logic. This formal knowledge representation is the basis of an integrated knowledge layer of a specialized software architecture being under development.**

***Keywords*—Knowledge representation, predicate logic, software requirements, user story.**

I. Introduction

The new software architecture with integrated knowledge layer developed by us will contain knowledge related to the software itself, i.e. behavioral and structural knowledge [1], [2]. These information will be gathered during the software development process.

To specify or analyze a software development process and to acquire relevant knowledge about it, we need to begin with the requirements and their representation.

Software requirements often represent a communication problem in software engineering[3]. Communication is needed between the customers and the developers of the software product. The success of the software development process depends on the amount of exchanged relevant and required quality information between the customers and software development team. The communicating units are on one side customers, end users, sometimes domain experts, business analytics, and economists. On the other side, the development team members can be found. If one side plays a highly dominant role in this communication, the most projects are lost and will not succeed. If it is the business side, functionalities and deadlines are defined regardless to the capabilities of the development team. If this communication is over-controlled by the development team, then a highly technical language will be used, which implies that the business side will be unable to gain necessary information about the project state and its tasks.

To eliminate the problems presented above and to deal with the problem to gather the right requirements, there are some well-known techniques such as the agile user stories[3].

Based on user stories, requirement model can be constructed that represents a more formal way of requirement expression using different classification criteria.

Using the requirement model, we will be able to describe them using predicates and terms of predicate logic.

The organization of the paper is as follows. Section II introduces user stories, Section III describes requirement model, and Section IV contains introduction into predicate logic. The proposed method of processing requirements from user stories to predicate logic is presented in Section V.

II. User Story

A. What Is a User Story?

User stories are used with agile software development methodologies as the basis for defining the functions a business system must provide, and to facilitate requirements management. User story is knowledge represented by one or more sentences in the everyday or business language of the end user that captures what a user does or needs to do as part of his or her job function.

User stories were introduced with extreme programming (XP), but can be adopted and used with any other agile development methodology.

A user story is an informal statement of the requirement as long as the correspondence of acceptance testing[4] procedures is lacking. Before a user story is to be implemented, an appropriate acceptance procedure must be written by the customer to ensure by testing or otherwise whether the goals of the user story have been fulfilled. User stories are written on *story cards*. When creating user stories, one of the developers (or the *product owner* in Scrum) gets together with a customer representative. The customer has the responsibility for formulating the user stories. The developer may use a series of questions to get the customer going, such as asking about the desirability of some particular functionality, but must take care not to dominate the idea-creation process. The user story focuses on the *who, what,* and *why* of a requirement in a simple, concise way, without trying to capture *how* it could be implemented.

In other words, user stories are a quick way of handling customer requirements without having to create formalized requirement documents and without performing administrative tasks related to maintaining them. The intention of the user story is to be able to respond faster and with less overhead to rapidly changing real-world requirements.

B. User Story Structure

User story consists of three parts (see Fig. 1):

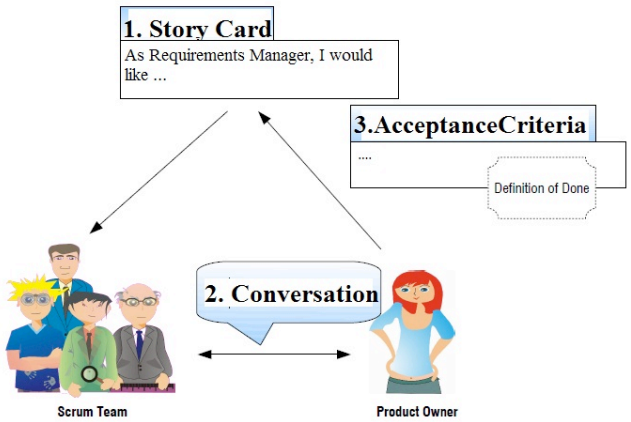


Fig. 1. User story structure[5].

1) story card (visible),
2) conversation (invisible), and
3) acceptance criteria (visible).

Story cards include usage-oriented assertions, which represent user requirements. These are incomplete, variable, and represent a promise for conversation.

Conversation is the source of information and knowledge. The closer is the realization, the more detailed and more complete it is.

Acceptance criteria introduce detail to the goals, make the user story more specific. These are variable and incomplete too, but must be finished at least before beginning with realization. Used to represent the *added value* of the user story.

The basic user story template is defined by [6] and [7] as a story card of the structure as follows in (1):

$$\text{As a} < role > \text{I want} < function > \text{so that} < value >. \quad (1)$$

User stories respectively story cards are being extended during conversations by usage scenarios that are used to explain and/or answer the *how* part of the story.

C. User Story Construction

In order to construct good user stories, Hartman [7] defines *INVEST*. INVEST is an acronym which encompasses the following concepts which make up a good user story:

***I*ndependent** Stories should be as independent as possible. In other words, stories can be worked on in any order.

***N*egotiable** A story is not a contract. A story *is* an invitation to a conversation. The story captures the essence of what is desired. The actual result needs to be the result of collaborative negotiation between the customer (or customer proxy like the product owner), developer and tester (at a minimum).

***V*aluable** If a story does not have discernible value it should not be done. The *so that* $< value >$ clause of the user story is there for a reason: it is the exact value we are trying to deliver by completing the story.

***E*stimable** A story has to be able to be estimated or sized so it can be properly prioritized. A value with high value but extremely lengthy development time may not be the highest priority item because of the length of time to develop it.

TABLE I
USER STORY EXAMPLE FOR A SIMPLE DVD STORE

As a	customer,
I want	to return the borrowed DVD to the store, when it is closed,
so that	I put it into the mailbox and do not need to wait until the store gets opened.
Scen. 1	I am a registered customer and I have a borrowed DVD. I come to the store, put the DVD into the mailbox, and leave. I assume that it will be found there, and later processed by the staff as I returned it personally during opening hours.
Scen. 2	I work for the DVD store. I check the mailbox. If there is any returned DVD, then I use the barcode reader to identify it, and I change its state to *returned* in our information system for the registered customer.

***S*mall** Obviously stories are small chunks of work. The size depends on the team and the methodology being used.

***T*estable** Every story needs to be testable in order to be *done*. In fact, testable means that acceptance criteria can be written immediately.

Table I presents an example user story including two usage scenarios.

D. Advantages and Disadvantages of User Stories

XP and other agile methodologies emphasize face-to-face communication between the development team and customers or customer proxy. User stories are the core documents during these conversations. Based on their features, there are some advantages and disadvantages to find.

Advantages:

- are very small, represent a small business value and can be implemented in a short period (days or weeks),
- allow changes in the requirements during the whole project life-cycle, indicated from both development team and customer,
- need a minimum of maintenance activities,
- maintain narrow contacts with customers (or customer proxy),
- allow splitting of processes into small steps,
- deal with incomplete and varying requirements, which are being refined in process iterations,
- ease development workload estimation, and
- require narrow contact with customers during the whole project life-cycle in order to implement the most important parts of the software first.

Disadvantages:

- are informal, so might need further concretization,
- might become hardly scalable for large projects, and
- are only possible starters of the conversation.

III. Requirement Model

The requirement model (RQM) is a detailed structured list of actions, which need to be implemented during development. It is a reference model that defines tasks and orients work activities of each member of the development team.

This model is aimed mostly for business analytics for business-related requirement specification, from which one the latter technical requirements are also created by software analytics and designers. To express enumerated alphanumerical information, graphical notations (i.e. diagrams) cannot be

TABLE II
REQUIREMENT MODEL EXAMPLE FOR A SIMPLE DVD STORE – OUR VERSION USING EXPLICIT KNOWLEDGE ABOUT COVERED USER STORIES

RID	Requirement description	Priority	Risk	Workload	ID of covered user story
1	(Non-functional) Technical requirements	1	Low	Low	all
1.1	Use Java technologies	1	Low	Low	all
1.2	Mailbox	1	High	Low	all
2	(Non-functional) Architectural requirements	1	Low	Low	all
2.1	Separated user interfaces	1	Low	Medium	aa123bb-1, ab234xy-3
2.2	Separated database	1	Low	Low	all
3	(Functional) Data-related requirements	1	Medium	High	all
3.1	Customer data – stored in a database	1	Low	Low	all
3.1.1	first (given) name	1	Low	Low	aa123bb-1
3.1.2	last (family) name	1	Low	Low	aa123bb-1
3.1.3	address	1	Low	Low	aa123bb-1
3.1.4	unique identifier	1	Low	Low	aa123bb-1
3.2	DVD information	1	Medium	High	all
3.2.1	unique DVD identifier – stored in a database, must be a number (barcode)	1	High	Medium	aa123bb-1
3.2.2	title	1	Medium	Low	aa123bb-1
3.2.3	type/category	2	Low	Low	aa123bb-1
3.2.4	title description	2	Low	Low	aa123bb-1
4	(Functional) Business-related requirements	1	Low	High	all
4.1	Staff functions	1	Low	Medium	all
4.1.1	checking the mailbox	2	Medium	Low	all
4.1.2	reading the barcode from the DVD using the reader	1	High	Medium	all

used. Therefore, there are used three types of tabular views instead of a graphical diagram. The three types are:

1) Requirement document view,
2) Traceability matrix view,
3) User allocation matrix view.

Requirement document view is a table for requirement hierarchy representation. Each row represents a requirement. The columns represent requirement properties such as:

- requirement ID,
- requirement description,
- priority,
- workload,
- risk of fail to implement the requirement.

Traceability matrix view displays the links between requirements and objects from other types of models, external files or other requirements.

User allocation matrix view displays the links between requirements and the users and groups who will fulfill them.

An example RQM document/traceability view is shown in Table II. It is the special case of an extended RQM developed by us to maintain backward traceability from requirements to user stories.

Each requirement is member of hierarchies of its superior and/or inferior requirements. A requirement must be precisely defined and clear to understand *before* assigning it to a developer or a team for implementation.

IV. PREDICATE LOGIC

Predicate logic is the generic term for symbolic formal systems like first-order logic, second-order logic, many-sorted logic, or infinitary logic.

This formal system is distinguished from other systems in that its formulas contain variables which can be quantified. Two common quantifiers are:

- existential qualifier (i.e. *there exists*) – $\exists$
- universal qualifier (i.e. *for all*) – $\forall$.

The variables could be elements in the universe under discussion (i.e. domain of discourse), or perhaps relations or functions over that universe. For instance, an existential quantifier over a function symbol would be interpreted as modifier *there is a function*. The foundations of predicate logic were developed independently by Gottlob Frege and Charles Peirce[8]. In this article, by the term *predicate logic* we will refer to *first-order logic*. Author of [9] considers the predicate calculus to be an axiomatized form of predicate logic.

A. Syntax and Semantics

Predicate logic[10] is defined using the following properties:

- The alphabet consisting of:
  - logical symbols ($\exists, \forall, \vee, \wedge, \rightarrow, \leftrightarrow, \neg$, parentheses, brackets, constants $\top, \bot$),
  - non-logical symbols (constants of the domain of discourse, function symbols ($f, g, \ldots$), predicate symbols (P_i^n)).
- Formation rules for:
  - terms, which can be *variables* or *functions*.
  - formulas, which can be
    - *predicates* defined by predicate symbol and terms,
    - *negations*,
    - *binary connectives* using the logical operator symbols over formulas, and
    - *quantifiers* using the quantifier symbols over variables and formulas.
- The definition of priority of operators, i.e. order or precedence.
- The definition of *free* and *bound* variables in formulas.

We present an example sentence using predicate logic alphabet in (2).

$$\forall x \forall y(P(f(x))) \rightarrow \neg(P(x) \rightarrow Q(f(y), x, z)) \qquad (2)$$

If f is a unary function symbol, P a unary predicate symbol, and Q a ternary predicate symbol, then (2) is a predicate logic formula.

B. Knowledge Representation Using Predicate Logic Formulas

Predicate logic can be used for knowledge representation in the form of *terms* that will represent knowledge about requirement identifiers, i.e. existence of the requirements – facts.

Predicates will be used to define relations between terms (e.g. requirements), and *functions* to derive consequences from an existing (hidden) relation or its nonexistence.

V. Proposed Method

Our method will use a modified RQM developed by us to maintain backward traceability from requirements to user stories (see Table II).

We propose the following method for knowledge representation:

1) gather the user stories, to each story a unique identifier string will be assigned,
2) analyze user stories and construct the modified requirement model, where the *Priority, Risk,* and *Workload* properties are gained as results of XP's planning game, and that contains the added column for identification of covered user stories,
3) transform each requirement of the extended RQM to a term of predicate logic,
4) transform each requirement group of the extended RQM to a function or predicate of predicate logic, representing initial relations between them.

The created knowledge base will be extended by new terms when creating an implementation class. New predicates might be included during implementation according to the created CRC (class-responsibility-collaborators) cards related to requirements.

VI. Conclusion

We presented our proposal of a methodology for knowledge representation about software requirements. The presented process begins in an agile way by user stories, then a modified requirement model is created to maintain backwards traceability from requirements to user stories. This step is also agile, because of usage of agile techniques such as multiple planning games. A formal step is used to finish the process: each requirement becomes a term of predicate logic, and if there is any relation between requirements, it becomes a predicate or function, respectively.

In fact, the decision whether to define predicates or functions has not been made yet. We need more experiments and studies to decide. Until further analysis, we do not consider the tool support for the method proposed.

Acknowledgment

This work was supported by the Cultural and Educational Grant Agency of the Slovak Republic, Project No. 050TUKE-4/2013: "Integration of Software Quality Processes in Software Engineering Curricula for Informatics Master Study Programme at Technical Universities – Proposal of the Structure and Realization of Selected Software Engineering Courses."

References

[1] O. Pločica and Z. Havlice, "Knowledge-based approach in information systems life cycle and information systems architecture," in *SAMI 2007, Proceedings, 5th Slovakian-Hungarian Joint Symposium on Applied Machine Intelligence and Informatics*, Poprad – Aquacity, 25.-26. 1. 2007, pp. 499–504.

[2] Z. Havlice, M. Paralič, V. Chladný, O. Pločica, H. Telepovská, Cs. Szabó, L. Samuelis, M. Varga, M. Beličák, J. Kunštár, O. Železník, and M. Révés, "Knowledge-based software life cycle and architectures," in *Computer Science and Technology Research Survey*. Košice: elfa, 2007, vol. 2, pp. 47–68.

[3] M. Cohn, *User Stories für die agile Software-Entwicklung mit SCRUM, XP u.a.* Verlagsgruppe Hüthig Jehle Rehm GmbH, Heidelberg, 2010.

[4] L. Samuelis, K. Frühauf, J. Ludewig, H. Sandmayr, and Cs. Szabó, *Software Testing Fundamentals: Introduction to Software Verification Theory*. Technical University of Košice, 2013.

[5] Ch. Fronia and H. Köppen. User Stories als Instrument des Requirements Engineering.

[6] S. Stamminger. Agile Requirements Engineering: User Stories, Schätzen, Planen.

[7] B. Hartman, "New to agile? INVEST in good user stories," May 2009. [Online]. Available: http://www.agileforall.com/2009/05/14/new-to-agile-invest-in-good-user-stories/

[8] A. G. Hamilton, *Logic for Mathematicians*. Cambridge UK: Cambridge University Press, 1978.

[9] A. A. Stolyar, *Introduction to Elementary Mathematical Logic*. NY: Dover Publications, Inc., 1970.

[10] D. Hilbert and W. Ackermann, *Grundzüge der theoretischen Logik*, 6th ed. Heidelberg, Germany: Springer, 1972.

Gesture Recognition using DTW and its Application Potential in Human-Centered Robotics

[1]Alena KOPANIČÁKOVÁ, [2]Mária VIRČÍKOVÁ (3rd year)
Supervisor: [3]Peter SINČÁK

[1,2,3]Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

[1]alena.kopanicakova@student.tuke.sk, [2]maria.vircikova@tuke.sk, [3]peter.sincak@tuke.sk

Abstract— **This article is focused on a Dynamic Time Warping (DTW) Gesture Recognizer–a program that creates a database of motions and then recognizes gestures using the depth data. To recognize a gesture, DTW warps a time sequence of joint positions to reference time sequences and produces a similarity value. We summarize the advantages and point on the current limitations of the system. Then we discuss the possible application potential as the gesture recognition is a technology commonly used in human-computer interaction applications. The focus is on the area of human-centered robotics, where the system must be able to interact with people such that the burden of adaptation lies with the machine and not with the human. We believe that transferring human communication, such as gestures to robots, creates various opportunities for more intuitive human-robot-interaction.**

***Keywords*—Dynamic Time Warping, human-centered robotics, gesture recognition, sensor Kinect, user interface.**

I. Introduction

The times, when computer was able to perform only basic operations like counting numbers are gone. Several years ago, nobody would say that the computer could be able to monitor and process human movements. Nowadays, the situation is different, as the research shows [1][4][5][6][7][8][9].

This article describes a system which uses depth information—invariant to color, texture and lighting objects, making it simpler to differentiate between the background and the foreground object. The first systems for depth estimation were expensive and difficult to manage in practice, but recently it has been published several works related to this topic because of the emergence of inexpensive structured light technology, reliable and robust to capture the depth information along with their corresponding synchronized RGB image. This technology has been developed by the PrimeSense company [11] and marketed by Microsoft XBox under the name of Kinect[2].

Using the mentioned sensor, Shotton et al. present in [7] one of the advances in the extraction of the human body pose from depth images, representing the body as a skeletal structure comprised by a set of joints. This article is focused on a system, using the basics of Shotton´s research, which is able to identify movements according to the motion patterns. The program is written by open source version in language C# and works with sensor Kinect, so that the user needs to have installed appropriate drivers (Kinect SDK)[2]. It has a broad prospect of application in computer games, robot control and home appliance control. The system is based on the using DTW algorithm, and we agree with [1] that it is a simple method to realize, and can meet the real time requirement.

The paper is divided into 4 parts. The part number two describes main functions of the program. In chapter 3, the reader can find details of DTW algorithm on which the system is based on. The principles of the algorithm are supported with the examples. The part 4 describes situations when the program is being used, its pros and cons and also some possible solutions to make the usage of the program easier. The last part defines the practical use of the program. Moreover, there are mentioned some examples from a real-life use, out of many possibilities, which the program offers.

II. Overview of the Program for Gesture Recognition

A. Capturing gestures

The program allows users to record their own gestures. We recommend starting recording approximately three seconds after pushing the button CAPTURE—at the moment when the recording starts. This time equals to the time needed to move about 1-1.5 meter from the sensor Kinect.

The best result can be achieved, if the gesture starts and ends at the exact time of recording—it means that movements must be faster or slower as obvious. The recording ends, when 32 pictures are recorded, and the program automatically switches to the reading mode.

B. Working with the library of saved gestures

If the user is satisfied with the gesture, it is possible to save it in —so that he can create his personal library of the gestures. If the recorded gesture is not suitable for user, he can easily record a new one. Moreover, it is possible for the user to create his own types of gestures. All he has to do is to add his own gestures to the combobox column in source code. The gestures that were previously recorded can be easily loaded and used again. The user just needs to set the location of the file in source code.

The proper working of a new recorded gesture must be tested. To do so, the program offers the function, which does not have to be recalled. The function runs when all the other functions are shut down. In case there is no match with one of the loaded gestures, the program displays UNKNOWN. If the match is found, there will be displayed NAME OF GESTURE.

Figure 1: From the left side: the output from RGB camera, the output from depth sensor and the points saved in the database as a vector of points in time.

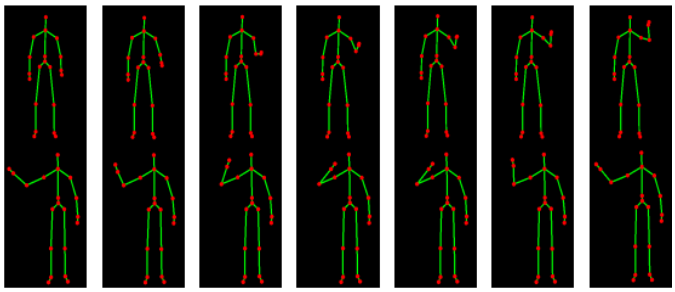


Figure 2: Two sample gestures: Right Hand Push Up and Left Hand Wave[1].

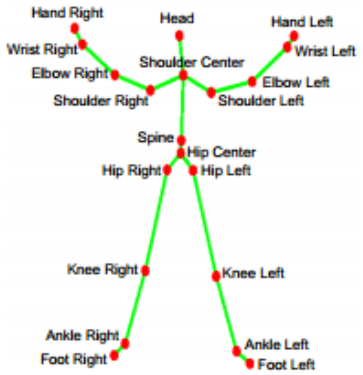


Figure 3: Kinect joints.

III. Principles of the DTW Algorithm

DTW measures the similarity among two sequences that may differ in time or speed. It can be applied for all types of data that can be represented linearly such as audio, video and graphics. It is well-known in speech recognition applications, e.g. [10].

The algorithm can be found in many applications of which the following two are the most common, according to [3].

The first one is a program that is able to identify a person according to the way he is walking. The algorithm works on a principle of detecting movement patters of a person from the video. Later, there is a video, where the walking is faster or slower. No matter how fast the person walks, the algorithm is able to identify the same person. The second one is an application that can recognize a voice. In this case, the algorithm must deal with different speeds of voice.

IV. Implementation of the DTW the Program for Gesture Recognition

DTW is used in Gesture Recognizer, which means that quality of recorded gestures does not depend on speed of recording but on the way of movement. That is the reason why the faster or slower movement does not affect the recorder gesture even though it is required for smooth run of a program.

A. Advantages of the system

The recording starts 3 second after the CAPTURE button is pushed, which equals to the time needed to move in front of the Kinect and to start making the motion. If 3 seconds are not enough for the end user, he can adjust the time needed for the transfer in source code.

The system senses all the joints of arms and trunk, so that the end user can easily record more gestures of the upper part of the body.

The program allows to create unique gestures and not only defined, which opens the limitless possibilities of movement of end users. All these gestures can be later stored in the gesture library.

Recorded gestures can be easily sent to the database on the internet, which allows more users to use the same database. Furthermore, it is possible to use any gestures and not only those that were recorder in the computer. The gestures were always saved with the same name and added date and time, which caused difficulties in the records. After the adjustment, it is possible to write any name into the textbox and after that the file is saved under the given name.

In order to create a new gesture, it was necessary to open the source code. Now, the situation has changed, and the new gesture can be recorded thanks to added textbox, where the name of the gesture can be added and the gesture is ready to be recorded.

B. Current limitations of the system

The system is able to process gestures only of one player. In case that two or more players want to record their movements, the program stops working and freeze.

Another disadvantage is that the program works only in 2D, which means that some types of gestures cannot be recorded. In the future, the program should work in 3D.

The last disadvantage is the manual control. In the future, Kinect should support voice control so that the end user would not have to move quickly in front of the Kinect and the system would not freeze.

V. Application Potential in Human-Centered Robotics

This part focuses on the potential of the system in the area of human-centered robotics, in which robots make their way into the everyday world that people inhabit. We see the application potential of the system in a great number of areas. Sechrist in [12] believes that "gesture recognition holds the promise of new frontiers in user interface that will change the face of coding forever, in much the same way that touch technology did a few years ago."

In general, we intend to use robot gestures as a non-verbal interactive modality that acts as an integral part for a given collaborative task in different human-robot scenarios. Nonverbal communication is one of the key issues in social robotics. Similar work was presented by [13] and [14]. Also, for example let's assume that the human waves with the left hand so that the robot recognizes this movement as greetings. It can respond to any gesture presented by the user, which is saved in the database.

The research topics of the human-centered robotics are presented in [16] (Figure 4.). The gesture recognition and the imitation of human movements can find its place in all of these partial topics. For example, also in the sense if a robot is

supposed to work in a human environment with the same objects and tools that humans use in everyday life, it may be beneficial to understand how humans perform motor skills and how the human skill repertoire is structured.

In [14] authors enumerate the following constraints to human robot interaction as a special case of human computer interaction:

1. The robot must be capable of real-time gesture recognition.
2. The system must be person independent.
3. The system must not require the user to wear specific clothes or inconvenient devices such as coloured markers or data gloves, since this can be tedious for the user.
4. The robot will face variable and possibly complex backgrounds against which a user operates it. A system requiring constant background is not flexible enough for real world applications.
5. The robot must cope with changing lighting situations. The requirement of constant lighting is a limitary restriction for any real world application.

The presented system fulfils all of the constraints. The future work will be to test the system in real-world environments. At present, we test the system incorporated in various projects, such us:

1. As a part of rehabilitation exercise in order to control the correct movements of patients.
2. Currently, we are setting up an experiment where the robot acts as a receptionist-it identifies people and serves them. The gesture recognition is a part of the overall system.
3. The field of ambient intelligence and smart houses- in this case each movement would be adjusted to some specific task that must be done; for example, raised right hand would mean that the light should be switched off.
4. As a contactless control of some programs, where the system and Kinect would control the programs. CD player can be considered as a good example, because the user would not have to press the button (for example play), but instead, he should just raise his hand.

Figure 5: The robot Nao shows children exercises and the DTW Gesture Recognizer monitors the movements.

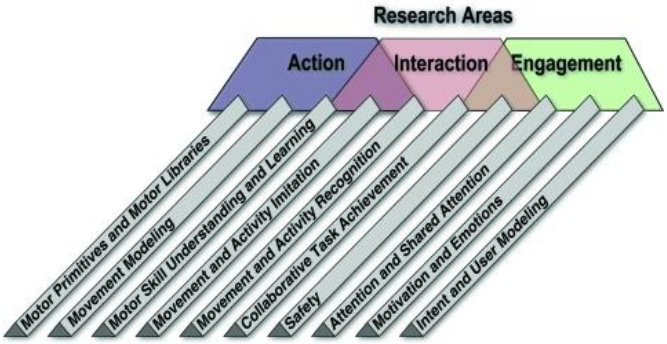


Figure 4: Attempt to structure research topics in human-centered robotics into three principal categories: (i) action, (ii) interaction, and (iii) engagement [16].

VI. Conclusion

This paper introduced the program DTW Gesture recognition. We illustrate the basic principles of the program and the algorithm that is based on DTW Gesture recognition. Based on real situations, we mentioned the pros and cons of the systems as well as improvements of the systems. In the future development of the program should focus on the direction of human-centered robotics, where the gesture body-motion recognition is essential in various areas.

References

[1] H. Duan and Y. Luo, "A Gestures Trajectory Recognition Method Based on DTW", in *Advances in Intelligent Systems Research*, ISSN 1951-6851, Proc. 2nd Int. Conf. on Comp. Science&Elec.Eng. 2013.

[2] Kinect DTW Documentation.[online]. [cit. 2013-02-26]. URL:<http://kinectdtw.codeplex.com/documentation>.

[3] Team of authors "Dynamic time warping". [online]. [cit. 2013-02-26]. URL:<http://en.wikipedia.org/wiki/Dynamic time warping>.

[4] K.Takahashi, S.Seki, R.Oka, "Spotting Recognition of Human Gestures from Motion Images," The Inst. of ElectromCS, Information and Comm, vol.36, No.7, pp.28-35, 1993.

[5] T. Wenjun, W. Chengdong, Z. Shuying and J. Li, „ Dynamic hand gesture recognition using motion trajectories and key frames", in *Advanced Computer Control, 2nd Int. Conf. on*, 3, pp163 –167, 2010.

[6] K. Lu, X. Li, "Gesture recognition research overview," in *Journal of Xi'an University of Arts & Science*, vol.9, No.2, pp.91-94, 2006.

[7] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake, „Real-time human pose recognition in parts from single depth images", in *CVPR*, 2011.

[8] C. Sminchisescu, A. Kanaujia, and D. Metaxas, „Conditional models for contextual human motion recognition", in *CVIU*, 104(2-3): pp. 210–220, 2006.

[9] V. V. Ganapathi, C. Plagemann, D. Koller, and S. Thrun, „Real time motion capture using a single time-of-flight camera", in *CVPR*, pp.755–762, 2010.

[10] T. B. Amin, and I. Mahmood, „Speech Recognition using Dynamic Time Warping", in *International Conference on Advances in Space Technologies*, pp. 74-49, 2008.

[11] PrimeSense Inc. Prime Sensor NITE 1.3 Algorithms notes, 2010.

[12] S. Sechrist, "Gesture Recognition Moves Beyond Gaming", in *Blog Post*, May 22, 2011 URL: <http://blog.smartbear.com/software-quality/bid/167282/Gesture-Recognition-Moves-Beyond-Gaming>

[13] A. Correa, R. Walter, L. Fletcher, J. Glass, S. Teller, and R. Davis, "Multimodal interaction with an autonomous forklift," in *International Conference on Human-Robot Interaction* (HRI), 2010 5th ACM/IEEE, Massachusetts Institute of Technology Cambridge, MA, USA, 2010.

[14] T. Ende, S. Haddadin, S. Parusel, T. Wusthoff, M. Hassenzahl, A. Albu-Schaffer, "A human-centered approach to robot gesture based commun. within collaborative working processes," in *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ Int. Conf.* , pp.3367-74, 2011

[15] J. Triesch and Ch. von der Malsburg, „Robotic Gesture Recognition", in: *Proceedings of the Bielefeld Gesture Workshop*, September 17-19 1997, Germany. Springer, LNAI 1371.

[16] S. Schaal, „The New Robotics-Towards Human-Centered Machines," in *HFSP J*. 2007 , pp. 115–126.

Hacking Language Learning

[1]*Vojtech RINÍK (1st year Ing.),* [2]*Ivan KLIMEK (3rd year PhD.)*
Supervisor: [3]*František JAKAB*

[1,2,3]Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

[1]vojtech.rinik@student.tuke.sk, [2]ivan.klimek@tuke.sk, [3]frantisek.jakab@tuke.sk

***Abstract*—This paper introduces a method for improving learning rate in the area of language studies. It is based on applying known methods for learning languages, particularly spaced repetition and usage of multimedia to help memorization. The novel part is functionality for extracting new words from user-selected content with the goal of increasing user's interest.**

***Index Terms*—Language learning, natural language processing**

I. Introduction

This paper starts with brief theoretical background about working memory based on work done by A. Baddeley. We use this knowledge to design a system for improving learning rate of second language based on spaced repetition.

Further sections introduce and give general overview of the proposed system, discuss the design of algorithms for selecting the right words to present to the user and deciding when to present a word again.

The final part describes the process of importing user-selected content and extracting and prioritizing words contained in the media.

II. Overview

From user perspective, the system consists of a single web application that allows users to perform basic tasks.

- Browse existing collection of books
- Upload their own media
- Learn vocabulary from selected chapter of selected book
- Practice user's existing vocabulary

In a typical learning session, user would first select a book they are about to read. They would pick a chapter, and the system would compare all words in that chapter with their existing vocabulary (which is collected over time).

As a result, the first few sessions might present mostly words that user already knows. For this purpose, a button would be available, which user could use to mark a word as already familiar.

If a word is new to user, the system would first show its definition, translation and other hints for memorizing. Then, as described in further sections, it would present some words again, multiple times, and the repetition would cause user to remember the word.

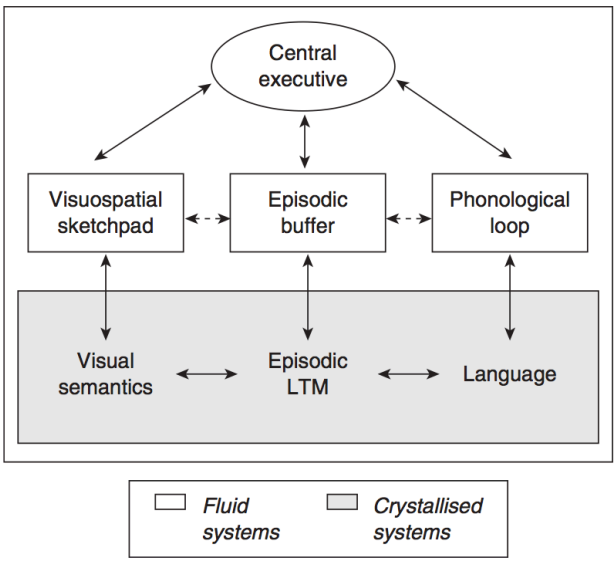


Fig. 1. Model of working memory.

III. Working memory

Our system is based on research made in area of working memory [1] [2] [3] and methods for learning new vocabulary. [4]

Figure 1 depicts a diagram of our working memory. The phonological loop is a time-limited system we use to remember short amounts of spoken information. The visuospatial sketchpad allows us to remember visual and spatial information. [5]

Episodic buffer stores "episodes" of information from multiple distinc memory subsystems. Baddeley (2007) describes it as:

> ...a temporary storage system that is able to combine information from the loop, the sketchpad, long-term memory, or indeed from perceptual input, into a coherent episode. (p. 148) [6]

Rehearsing items in the short-term memory strengthens their associations in the long-term memory. [7], [8]

The system is based on repetition and the concepts discussed in this paper are related to memorization via repetition.

TABLE I
WORDS DATABASE

Word	No./attempts	Last attempt	Confidence	Ideal delay
das Fahrrad	4	12:34	0.4	5m
der Bürgersteig	5	12:35	0.45	3h
steigen	0	✗	0	✗

IV. SELECTING WORDS

The main entry point of our system is where user clicks "start learning" and the system selects a word for them.

At this point we introduce the main data structure used to store information about user's progress, the table of words. (table I)

Column **confidence** indicates how confident we are that user has stored the word in their long-term memory. It takes values between 0 and 1, but does not necessarily mean the probability of knowing the word.

Column **ideal delay** indicates that there is an ideal delay for the word to be presented to user again.

Before we can proceed to selecting a word, there are a few important facts we have to consider. The collection of all words to learn may be large, and user is able to learn only certain amount of words at a time.

Listing 1 shows a pseudo-code of the algorithm that we use for selecting the next word.

```
buffer_size = 10
existing_words = Word
  .where('confidence > 0.0 AND confidence < 0.8')
if len(existing_words) < 10:
  to_add = buffer_size - existing_words
  words = existing_words + Word.random().limit(to_add)
else:
  words = existing_words
return words
  .order_by('ABS((last_attempt + ideal_delay) \
            - DATE()) ASC')
```

Listing 1: Pseudo-code for selecting the next word

First we try to find a list of such words that the user is currently learning, the ones they have in their mind at the moment. If there are not enough such words, we select some random words from the list of all words. To order the result, we compute how many seconds are we away from the ideal time for their repetitions, and put the ones closest to the ideal time at the beginning.

V. SPACED REPETITION ALGORITHM

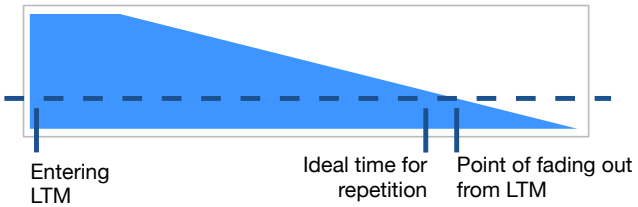


Fig. 2. Repeating in long-term memory

The main goal of our system is to get a word from short-term, temporary memory, to the long-term memory. Figure 2 shows how the strength of word association changes after entering the long-term memory. At first, the association is strong, but eventually it weakens.

The design goal of the algorithm is to offer user repetition of word just before getting to the point when user cannot remember it anymore. Offering the word earlier than that would not be optimal - if the user can remember word easily, than it is too early to repeat it.

Offering the word after the point of forgetting would have obvious consenquences. But offering the word earlier or later than optimal is never necessarily negative - repetition strengthens association in any case.

This gives foundation to our algorithm - we start by offering words with some predefined delay (based on historical data or sensible default values), with respect to the number of previous repetitions and difficulty of the word.

After starting, we record data in a table-like structure as depicted in table II. For better understanding, figure 3 shows a visualization of this table.

TABLE II
EXAMPLE DATA COLLECTED WHILE INTERACTING WITH THE PROGRAM

Word	Number of attempt	Delay	Success
das Fahrrad	1	0	✗
das Fahrrad	2	10s	✗
das Fahrrad	3	30s	✓
das Fahrrad	4	5m	✗
der Bürgersteig	1	0	✗
der Bürgersteig	2	10s	✗
der Bürgersteig	3	30s	✓
der Bürgersteig	4	2m	✓

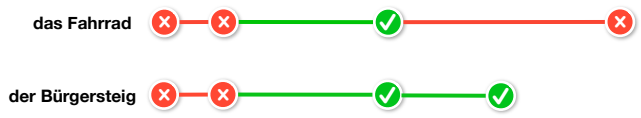


Fig. 3. Visualization of table of attempt data

Our intention is to figure out the ideal delay between third and fourth attempt for a new word, based on previous data about user. (For first few attempts we use fixed length as they are in most cases failures.) This is depicted in fig. 4.

As for general overview, we could say that if user is doing too good, we are giving him too little time between repetitions. In opposite case, if user is permanently failing, there might be too much space. Either way, we have to adjust the delay for new word.

The pseudocode for our algorithm might look similar to the following snippet in the listing 2.

Our example has only two previous words, each with with different delay before the fourth attempt and different results. The algorithm would see that the success to failure rate is about the same, and it would set the delay for the next word to the same value as average of the last two words. Looking back at table II it is obvious that the new delay will be set to $\frac{2+5}{2} = 4$ minutes.

This concludes the algorithm for finding the optimal delay for the next word in the queue.

```
adjustBy = 0.2
attempts = Attempt
  .where(attempt_number=4)
  .order_by(created_at=desc)
  .limit(10)
averageDelay = attempts.average('delay')
successful = attempts
  .where(result='success').count()
failed = attempts
  .where(result='failure').count()

if (successful > failed):
  newDelay = averageDelay * (1 + adjustBy)
else if (successful < failed):
  newDelay = averageDelay * (1 - adjustBy)
else:
  newDelay = averageDelay
```

Listing 2: Pseudo-code for computing the new delay

Fig. 4. Visualization of the deciding process for new word

VI. Confidence level

As mentioned in section IV, after every repetition, word is assigned a new confidence level. The formula for deriving confidence is simple: Success increases confidence, failure decreases confidence. Higher number of attempt causes greater change in confidence. In plain speak, if user failed at the beginning of learning, it does not change the confidence much, but if they failed after tenth repetition, the confidence can be decreased significantly.

```
change = (1 if result == 'success' else -1) * \
         (number/max_number)
confidence = confidence * (1 + change)
confidence = 1 if confidence > 1 else confidence
```

Listing 3: Pseudo-code for assigning new confidence

VII. Presentation of words

The system incorporates various resources to provide different sensations as a way to maximize use of the episodic buffer.

Translation of the word is the basic association.

Related image engages the geospatial short-term memory and creates an association between a word and an image might help to remember it.

"Mem" is a funny or witty play with the language to help memorization of word. These would be submitted by users.

Audio adds another new kind of experience. The recordings could be user-submitted, or generated by a synthesizer.

VIII. Connecting with content

Figure 5 describes the purpose of importing user-selected content into the system. The idea is to extend user's collection of currently studied words by new words contained in imported media.

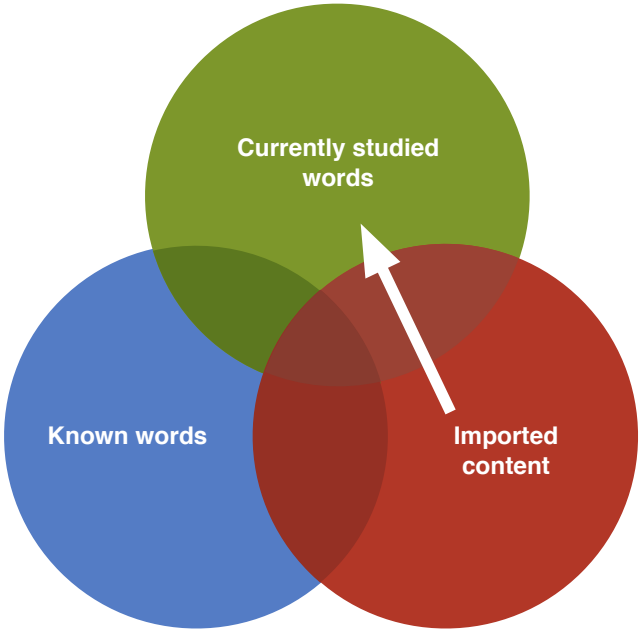


Fig. 5. Connecting with content

This process is what makes our proposed system different from all available tools used for learning vocabulary. Existing tools can be in general divided in two categories:

- Systems with fixed collection of words
- Systems with custom content supplied by users

Systems with fixed collection of words are usually curated by language teachers, they select the right words and choose their priority for learning. The other category allows users to write down a list of words they would like to memorize, and the system helps them with that.

Our proposed system is based on the first category, except the collection of words is not fixed, but automatically extracted from media selected by user.

Some resources are required before we can start importing and comparing user-selected data.

Dictionaries are used to find if the input word is a part of language, and to find some meta-information about the word.

Linguistic corpora are used to find word frequencies in the language.

The process of importing words consists of comparing with user's word database (to prevent duplicates), then dictionary lookup, and consenquently frequency lookup. Words are ordered by priority which is computed from frequency and possibly other rules, and inserted into user's database for learning.

To prevent learning words that user has already mastered, but they were never in system's database, the user interface provides an option to mark word as "already in long-term memory".

IX. Conclusion

During our research we implemented the prototype of described system and conducted basic testing on a few targets. It is yet to be determined how effective it is compared to other learning methods.

ACKNOWLEDGMENT

This work is the result of the Project implementation: Competency Centre for Knowledge technologies applied in Innovation of Production Systems in Industry and Services, ITMS: 26220220155, supported by the Research & Development Operational Programme funded by the ERDF.

REFERENCES

[1] A. Baddeley, *Working Memory*, ser. Oxford Psychology Series. Clarendon Press, 1986, no. 11.

[2] A.Baddeley and G.Hitch, "Working memory," *The psychology of learning and motivation*, vol. 8, 1974.

[3] A.Baddeley, "The episodic buffer: A new component of working memory?" *Trends in Cognitive Science*, vol. 4, pp. 417–423, 2000.

[4] I.Rock, "The role of repetition in associative learning," *The American Journal of Psychology*, vol. 70, pp. 186–193, 1957.

[5] L.Henry, *The Development of Working Memory in Children*, ser. Discoveries & Explanations in Child Development. SAGE Publications Ltd, 2012.

[6] A.Baddeley, *Working Memory, Thought, and Action*, ser. Oxford Psychology Series. Oxford University Press, USA, 2007.

[7] A.Miller, "The magical number seven, plus or minus two some limits on our capacity for processing information," *Psychological Review*, vol. 101, pp. 81–97, 1956.

[8] R.Atkinson and R.Shiffrin, "Chapter: Human memory: A proposed system and its control processes," *The psychology of learning and motivation*, vol. 2, pp. 89–195, 1968.

How SCYR Proceedings are Created

[1]*Milan NOSÁĽ (2nd year),* [2]*Emília PIETRIKOVÁ (3rd year)*
Supervisors: [1]*Jaroslav PORUBÄN,* [2]*Ján KOLLÁR*

[1,2]Department of Computers and Informatics, FEI TU of Košice, Slovak Republic

[1]milan.nosal@tuke.sk, [2]emilia.pietrikova@tuke.sk

***Abstract*—This paper examines perceptions of automated proceedings creation, resulting from implementation of the ProcCr application for the purposes of the SCYR conference. After ten years of manual creation of conference proceedings by our predecessors, we decided automate this work in order to simplify the whole process as well as to shorten the load time. The paper describes not only paper sorting or merging, but particular attention is paid to paper parsing, and to generation of table of contents and author index. The aim of this publication is also to show other participants that the SCYR conference progresses each year. Along with other changes performed by the conference organization committee, this work represents a significant contribution to increasing quality of the Scientific Conference of Young Researchers at Technical University of Košice.**

***Keywords*—Proceedings, generative programming, DSL, Java, PDF, TEX.**

I. Introduction

As most people from the university environment, we often participate at various conferences. Usually, we obtain a paper block with a pen, or a small gift (e.g. a calendar), but certainly we obtain conference proceedings on USB, CD or it is available online. But have you ever thought about how your proceedings were created?

The use of appropriate tools allow easy creation of the proceedings, further editing, indexing and the possible touch up of submissions. Management systems can assist in reducing administrative burdens and reduce time demands on proceedings' editors [1].

One of the popular proceedings management systems might be ProceedingsBuilder which is a hybrid of a web content management system and a workflow application, serving as paper collector, sorting papers by titles or author names, with possibilities of web-based proceedings creation [2]. Other well-known systems are large conference management systems like EasyChair [3] or OpenConf [4] which are widely used to manage academic conferences, also providing a few not so perfect abilities to create, or rather pre-create conference proceedings.

Apart from requiring various fees, not every management system should satisfy particular editor's requirements. This situation is sometimes complicated by the fact that each conference has its traditions along with a specific format of its proceedings. Moreover, these systems often do not foresee specific activities necessary to build a conference proceedings, such as paper parsing, or generation of table of contents and author index.

In this paper we present ProcCr – a simple conference proceedings application with PDF-assembly like features. We also present architecture and development of this prototype with an approach using generative programming [5]. With ProcCr, a proceedings editor will be able to manage common activities associated with conference proceedings creation. The paper closes with the prototype evaluation and a look at the future perspectives.

II. Motivation

Ordinarily, conference proceedings may be accessible as web page, book, or PDF document. The simpler way of how to create conference proceedings is to create a web page, while the proceedings creator needs to sort collected papers and to generate links to the papers. This can be done manually, what can be a tiresome work but not so time consuming. However, if done automatically, it can take only a few seconds to a few minutes. Automation of such work would include implementation of a script generating html code with sorted conference papers and link to these papers, so it is possible to find particular papers only by finding them in the proceedings web interface and one other click.

Second, more common form represents a book or a PDF document, which might be challenging for the proceedings creator, as his (or her) work consists of more, mostly interdependent parts. This includes Table of Contents generation, Author Index generation, page numbering, and generation of other pages, usually containing also number of pages and other summary information, such as impressum and masthead. If done manually, it can take a lot of time, and moreover, it is very prone to mistakes, as it is human who sorts the papers and counts page offsets. However, if done automatically, proceedings generation can take only a few minutes, but most of all, the resulting document does not contain any mistakes, so it does not require any additional corrections.

Despite wide range of possibilities, we decided to make our own application ProcCr, which results in a PDF document of the SCYR conference proceedings. PDF document is a traditional form of the SCYR proceedings, what is also related to the fact that the conference paper templates are in two forms: MS Word and LaTeX. Moreover, our intention was not to perform radical changes, but only to automate otherwise manual work and to simplify the whole process for the future generations.

III. First Attempts

Our first attempts relate to an effort to create the entire document in LaTeX, what has proved to be insufficient as the conference papers are collected as PDF documents and there

was still a need to manage various scripts and other support applications.

The integration of applications developed in systems programming languages coordinated through scripting languages has been common practice for a long time now. Shell scripting in Unix systems is probably the most notable example, where constructs such as pipes [6] (which connect the output of a process to the input of another one) allow one to perform tasks combining series of programs implemented in different languages, or even other scripts.

Therefore, our further attempts were focused on a diploma thesis [7] implemented in the Python programming language. The work was focused on a module called pyPdf which was used to merge PDF documents together, rotate pages, split and crop pages, and decrypt or encrypt particular PDF documents. Graphical user interface for generation of the conference procedings within the mentioned experiments used libraries wxPython, Tkinter, and PyGTK [8].

However, despite a good design (see Fig. 1), the resulting application still required a lot of manual work, such as entering the names of authors and the titles of papers[1]. Moreover, it could not be really applied to proceedings creation as it contained a few serious errors related to document coding.

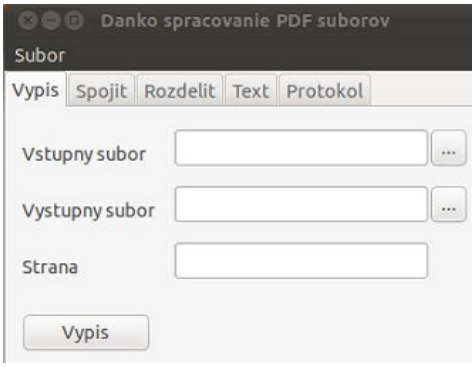


Fig. 1: Graphical application as one of the first attempts

Correction of these errors was evaluated as inefficient, thus we decided to develop our own application *ProcCr*, partially alleging the results of our first attempts. We considered wide range of document-creation possibilities, for instance lightweight markup languages allowing to write structured plain text format and convert it to structurally valid documents, such as Markdown or AsciiDoc, as well sa DocBook, a semantic markup language for technical documentation [9]. And finally, we combined TEX, simple DSL (Domain Specific Language) [10] [11], Java, and pdftk application.

Although it does not represent the most appropriate way of the proceedings creation, ProcCr is a mixture of tools and approaches we use daily and we are familiar with, and thus we were able to implement it in a very short time with plausible results.

IV. ProcCr Architecture

Since solutions performed directly from Java proved to be inefficient and insufficient, ProcCr uses the existing application pdftk [12] for PDF overlaying. Thus, control of the entire process lies in a script calling the Java program along with other external programs. Given this fact, ProcCr only works with command-line user interface, and supports the following parameters:

`-ptb` — Relative path to PDF document representing a blank page,
`-intro` — Number of pages in proceedings introduction pages, e.g. Sponsors, Preface etc.
`-title` — Number of pages in proceedings title,
`-end` — Number of pages in proceedings final segment,
`-cont` — Number of pages in proceedings Table of Contents — Number of these pages as well as number of Author Index pages are not known in advance, and thus it is necessary to run the script at least twice: First to generate Table of Contents and Author Index, second to get number of the pages.
`-index` — Number of pages in proceedings Author Index,
`-titleP` — Relative path to PDF with proceedings title page,
`-introP` — Relative path to PDF with proceedings introduction pages,
`-endP` — Relative path to PDF with proceedings final segment pages,
`-input` — Input DSL file with information about proceedings papers,
`-output` — Output TEXfile to which Table of Contents, Author Index, and paging are generated,
`-combined` — Relative path to PDF with all papers, including title page, introduction pages, empty pages reserved for Table of Contents and Author Index, and final segment pages — This document is generated by composition of other program parameters.
`-sort` — Boolean flag indicating whether the papers should be sorted or not — If sorted, papers are listed in the order of conference categories and names, and if not, the absolute order (from the DSL file) is used.

Figure 2 depicts a process of how our Java program ProcCr works, dividing it in two stages: *Proceedings paper assembling* and *Table of Contents and Author Index generation*.

A. Proceedings paper assembling

Within this stage, ProcCr first reads DSL file from which it creates model of papers. These papers are then sorted if appropriate, and merged together with the title page, Sponsors page, masthead, Preface as well as with impressum, all into one PDF document without page numbering, containing blank pages reserved for Table of Contents and Author Index (generated during the following stage). For each paper, page numbers are counted synchronously.

B. Table of Contents and Author Index generation

Next, Author Index is generated from the sorted list of papers with page numbers. Both are subsequently used to generation of TEX document which contains completed Table of Contents and Author Index and other pages reserved for the rest of the proceedings. Thus the output consists of two documents, one with proceedings papers and the second with

[1]Meanings of particular Slovak words in the application screenshot: `Danko: Student's name, Spracovanie suborov: File Processing, Subor: File, Vypis: Printout, Spojit: Merge, Rozdelit: Split, Text: Text, Protokol: Protocol, Vstupny subor: Input File, Vystupny subor: Output File, Strana: Page`

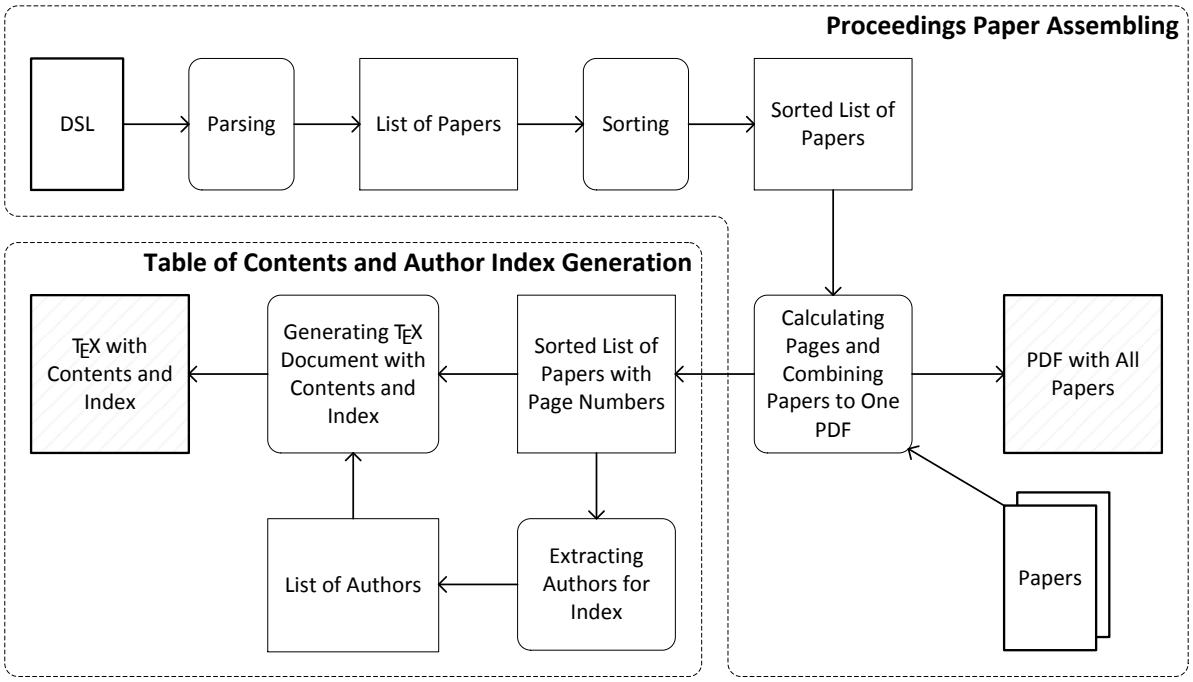


Fig. 2: ProcCr application process architecture

Table of Contents and Author Index. For PDF assembling, *itext* library is used.

After these two documents are generated, it is necessary to call a script to build the TEX document, providing PDF document of the same length (number of pages) as the document containing proceedings papers. In order to create final conference proceedings document, these two PDFs need to be overlayed into one document, what is ultimately handled by the pdftk tool.

In this process, the intermediate step using TEX document is very important as it significantly simplifies generation of the document containing Table of Contents and Author Index. It is possible to do this step manually in Java with various PDF libraries, however, after several attempts this proved to be incomparably more difficult.

When generating TEX document, which is generated from the DSL file data, it is important to mention accented characters, such as *l'* or *t'*, that need to be handled separately. That is, after TEX generation, postprocessing is performed, substituting each of these characters by appropriate TEX code sequences.

Another way to solve this issue could be carried out using `inputenc` package for `utf8` to generate the document with the `utf8` encoding. However, this encoding is not supported by every TEX environment (mostly of Windows OS, e.g. TexnicCenter). The method, which is actually applied, exempts from encoding manipulation as it generates `ascii` characters directly.

Next, Author Index generation is performed. Basically, this process is simple, assigning a key to each author by name through a hashtable. Next, page numbers of particular papers are assigned to corresponding keys.

C. DSL

DSL is a simple domain-specific language, the structure of which is as follows:

```
IT
Adrián PEKÁR, Marek DUFALA
Contribution to Real-Time Network Traffic
KPI
4
poster
```

And the meaning of each line is as follows:

1) Section to which the paper belongs (IT/EEE)
2) Author names separated by commas
3) Paper title
4) Department
5) Number of pages
6) Presentation method (poster/oral), important for sorting

Paper title and section are important data, since ProcCr seeks particular papers on the relative path `{section}/{title}.pdf`.

V. Results

The resulting proceedings is a PDF document containing Table of Contents with cross references directly to pages of particular papers, and Author Index with the cross references as well.

This method of conference proceedings generation offers several advantages, for instance, it significantly reduces the time load. Compared to the previous years, when the entire process consumed approximately one day of work with Adobe Reader by our predecessors, last year (2012) the process took up to ten minutes. Formerly, the biggest problem was manual calculation of page numbers and offsets for Table of Contents and Author Index along with manual creation of particular cross references.

Moreover, a significant benefit is also an additional completion of proceedings in case any paper got "lost" during the registration phase. In the past, serveral conference papers were

not published after they got lost, since they were "found" after the proceedings had been already completed.

Naturally, there is still much room to improve: Introduction and final segments of the proceedings still have to be created manually as they require periodic update of scientific and conference committee, Preface, ISBN, and other data each year.

Moreover, the application is used on client side, requiring that the user's (proceedings editor's) system meet several requirements, e.g. TEX interpreter, pdftk tool, and Java. Further, the script is currently applicable only MS Windows systems, hence it currently cannot be used cross-platform.

VI. Conclusion and Future Work

This paper examined perceptions of one of the invisible activities, usually unnoticed by conference participants, resulting from our own experience with generation of the SCYR conference proceedings. Although proceedings creation seems to be fairly burdensome, if the editor is able to find a way of how to reduce the load, this job might become quite fast and pleasant.

Practical contribution of this work lies in revealing new possibilities for automation of otherwise manual work related to creation of large PDF documents compound of several smaller documents, with a need for sorting and generation of table of contents (including cross reference) or author index. We would also like to draw your attention to the use of a simple domain-specific language [11] as a modern way to simplify specifically focused issues.

After setting up changes described in this article, the time load was markedly reduced not only in the proceedings creation stage, but also in the preparation stage. Compared to other formats, PDF format is more challenging to generate, which is currently handled faster than the one in HTML format prepared by our colleagues at Czech Technical University in Prague. We are able to create the proceedings up to a few minutes, while they usually do the proceedings approximately two days.

Based on the e-interview we made with them, we are not the only ones who would appreciate automation of this work. Moreover, even they did not admit any problems, we have already experienced a few broken links in their proceedings, showing that neither proceedings in HTML format is not that simple as it seems to be.

In the future work, we also plan to deal with other issues that are not yet managed. Since other innovations are prepared for the SCYR 2013 including registration through a conference management system, we plan to use its data about authors and papers directly. Further, we plan to shift from client side to server side. That is, proceedings creation will be initialized directly from the conference management system, thereby minimizing editors' work who will not have to install required tools and applications. This way, besides other necessary work such as ISBN registration, editors will only have to formally check whether the proceedings was generated correctly.

Other of our visions include proceedings extension by the HTML version as part of the conference website which, for example, will simplify library registration of participants' articles.

The inventive activities described within this study are seen to satisfy the automation requirements for the conference proceedings creation. Together with the conference management system, they are seen as an opportunity to strengthen and increase the SCYR conference qualities. This study can also offer future conference organizers and proceedings editors the benefit of experience and lessons learned on the skill of creating conference proceedings.

Appendix
E-Interview with Other Proceedings Editors

As part of this study, we made a short e-interview with our colleagues Petr Němeček and Libor Husník from Czech Technical University in Prague, editors of the Poster conference proceedings. Our aim was to compare time requirements and problems related to proceedings creation of the similarly oriented conference and of the different format.

- **For the Poster conference you publish proceedings in HTML format. Why did you decide right for HTML?**
  Orientation in the menu is clear and quick. This format is common and can be used anywhere.
- **How much time does it usually take to create the proceedings?**
  It usually takes up to two days.
- **Have you encountered any problems in this work?**
  No, not when creating the proceedings.
- **Do you use any mechanism for automation of your work?**
  No, we do not use any. Automated linking of paper links would be helpful and useful.

Acknowledgment

This work was supported by project VEGA 1/0341/13 Principles and methods of automated abstrac- tion of computer languages and software development based on the semantic enrichment caused by communication.

We would also like to thank our colleagues Petr Němeček and Libor Husník from Czech Technical University in Prague for a short e-interview for the purposes of this study.

References

[1] E. Brazil, "Creating conference proceedings: Tips and tricks from the trenches," Tech. Rep., 2002. [Online]. Available: http://richie.idc.ul.ie/eoin/research/UL-IDC-02-03.pdf
[2] J. A. Mülle, K. Böhm, N. Röper, and T. Sünder, "Building conference proceedings requires adaptable workflow and content management," in *Proceedings of the 32nd international conference on Very large data bases*, ser. VLDB '06. VLDB Endowment, 2006, pp. 1129–1139.
[3] H. Qunoo and M. Ryan, "Modelling dynamic access control policies for web-based collaborative systems," in *Data and Applications Security and Privacy XXIV*. Springer Berlin Heidelberg, 2010, vol. 6166, pp. 295–302.
[4] A. Pesenhofer, R. Mayer, and A. Rauber, "Improving scientific conferences by enhancing conference management systems with information mining capabilities," in *First International Conference on Digital Information Management*. IEEE, 2006, pp. 359–366.
[5] K. Czarnecki, K. Østerbye, and M. Völter, "Generative programming," in *Object-Oriented Technology ECOOP 2002 Workshop Reader*. Springer, 2002, pp. 15–29.
[6] E. Quigley, *UNIX Shells by Example*, 4th ed. Prentice Hall, 2004.
[7] P. Daňko, "Functional approach to scripting (diploma thesis)," 2012.
[8] G. Polo, "Pygtk, pyqt, tkinter and wxpython comparison," *The Python Papers*, vol. 3, no. 1, pp. 26–37, 2008.
[9] B. Stayton, *DocBook XSL: The Complete Guide*, 4th ed. ACM, Sagehill Enterprises, 2007.
[10] M. Fowler, *Domain-Specific Languages*. Addison-Wesley Professional, 2010.
[11] J. Kollár, J. Porubän, and S. Chodarev, *Modeling and Generation of Software Architectures*. Elfa sro, 2012.
[12] S. Steward, "pdftk–the pdf toolkit," Tech. Rep., 2006.

Human Activity Recognition using Mobile Devices

[1]*Marek Novák (4^{th} year),* [2]*Patrik Mucha (2^{nd} year master)*
Supervisor: [3]*František Jakab*

[1,2,3]Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

[1]marek.novak@tuke.sk, [2]patrik.mucha@tuke.sk

***Abstract*—In this paper we present a model for human activity recognition from accelerometer sensor on a mobile device. Our model can recognize three human activities: sitting, staying, walking and a state when mobile phone is lying on a table. The recognition is based on a heuristic decision tree classification method learned from a data of 20 users. The heuristic method uses mean, variance and vector magnitude of an accelerometer vector. The experiments were realized on an Android application with 25 respondents with promising accuracy above 92%. We outline also the idea how activity recognition alongs with time of an activity and duration of an activity can be used for users' behavior anomaly detection. Thus we can detect unusual long activity, unusual activity, changes in daily rythms, etc.**

***Keywords*—accelerometer, activity recognition, anomaly detection, decision tree, heuristic**

I. Introduction

Research and development in the area of *Ambient Assisted Living* (AAL) is highly actual nowadays. It is fostered by an assumption of a high number of elderly and disabled people in their 80s in the half of 21^{st} century [1], which may have economical and social consequences. There is a fruitful progress in this research area, however many approaches, especially when we are speaking about proactive services, are still very far from real life application.

In paper [2] we have presented a model for finding anomalies in user's presence patterns. Anomaly detection was based on three values: start time of an activity, duration of an activity (either in active or inactive state) and room location. Thus we were able to classify unusually long periods of inactivity and unusual presence. Input data were collected unobtrusively from passive infra-red sensors (PIR) or bed/sofa pressure sensors. In this paper we continue in the previous work from the point of anomaly detection, however we changed firings from ambient motion sensors to actions detected on mobile phones.

The majority of currently available smart phones on a market contain tri-axial accelerometers that measure acceleration in all three dimensions. These accelerometers are also capable of detecting the orientation of the device (helped by the fact that they can detect the direction of Earth's gravity), which can provide useful information for activity recognition. Accelerometers were initially included in these devices to support advanced game play and to enable automatic screen rotation but they clearly have many other applications.

We present an algorithm for detection four different actions of a user: standing, walking, stationery (sitting, lying) and state when mobile phone is put on a table. For action recognition we have created a heuristic decision tree model based on a features as: mean, variance of accelerometer vector and vector magnitude.

Section II. presents other works devoted to action or activity recognition by accelerometers. Section III. presents our model based on heuristic decision tree classification algorithm, section IV. discusses experimental study and section V. outlines how activity recognition is used for human behavior anomaly detection.

II. Related Work

The idea of activity recognition by accelerometer is not absolutely new. The majority of earliest work in accelerometer based activity recognition focused on the use of multiple accelerometers placed on several parts of the user's body.

In one of the first works L. Bao and S. S. Intille [3] proposed a solution to classify 20 different activities were subjects wore 5 biaxial accelerometers. They acquired accuracy ranging from 70% to 96%. Only some activities like riding an elevator or stretching had the accuracy below 50%. Tapia et al. [4] used three accelerometers and heart rate monitor to recognize tree activities: walking, cycling and rowing in different intensities. B. Kaluža and M. Gams [5] used 4 tags attached to a person to recognize 3 different postures: lying, sitting, walking. While monitoring a person for a longer time they created a spatial activity matrix to search for variances in users' living habits.

Although the higher number of sensors provide more valuable data and thus the recognition accuracy is higher, such systems are unacceptable to users and one can hardly expect that somebody would wear four or five accelerometers while doing standard daily activities.

With the occurence of smart phones, there have already been done some research to recognize activities by accelerometer sensor. In works [6] [7] the authors focused on dynamic activities like walking, jogging, going upstairs, going downstairs. Since we want to monitor a user in his living environment from a long term perspective we want to monitor activities like: walking, lying, sitting and standing and use them for daily living anomaly detection. Especially in our environment people are usually living in one floor flats or houses, therefore we are not interested in stairs climbing. Moreover, even if considered, these activities takes only a small portion of time during the day.

Just recently appeared projects like *Fitbit* [8] offering a wristlet that can measure steps taken, calories burned or the length of sleep. Our algorithm can be used in a similar manner for an elderly to monitor duration of daily activities and the time they are realized. Thus after some learning period we can

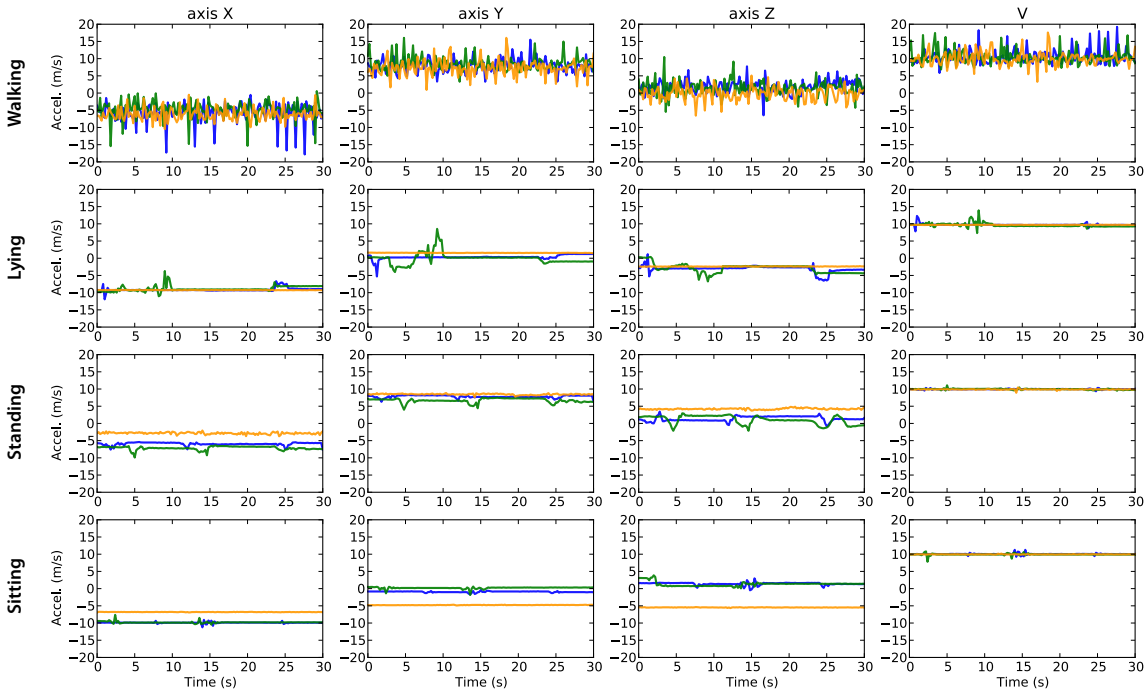


Fig. 1. Measured values from accelerometer for three different users (blue color is user1, green color user2 and orange color user3) and different actions: walking, lying, standing, sitting.

detect an unusual behavior which may indicate some health problem.

III. Activity Recognition

A. Accelerometer's Data

An acceleration sensor in a smart phone measures the acceleration applied to the device, including the force of gravity [9]. Acceleration of a device is formulated as (1), where $\sum F$ is the sum of forces applied to the sensor itself, m is phone's mass and g is gravity acceleration of a magnitude $9.81m/s^2$. Since gravity is included, the accelerometer reads value -9.81 in z axis, when phone is lying on a table.

$$A = -g - \left(\sum F/m\right) \tag{1}$$

In Fig.1 there are depicted measurements from accelerometer for three different users. The measurements were realized for four different actions: *walking*, *lying*, *standing* and *sitting* (shown in rows) in a time interval of 30 seconds. Measured values are in three axes: *x*, *y*, *z* and the last column *v* represents an acceleration vector magnitude, given by formula (2). The users wore the smart phone in their pocket with *y* axis heading to the sky as depicted in Fig.2.

$$v = \sqrt{x^2 + y^2 + z^2} \tag{2}$$

Our aim was to measure values in real conditions, therefore there is a peak for a green user while lying (he wriggled on a bed) or an orange user doesn't put the mobile phone in a precise position, which is visible in shift of values, especially while sitting.

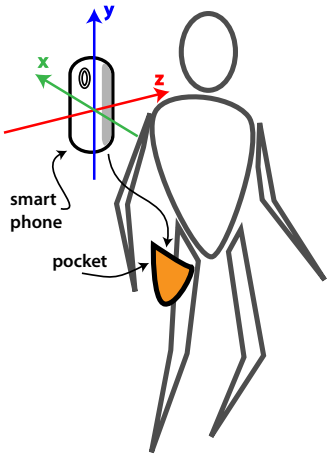


Fig. 2. The orientation of a smart phone with regard to the user.

B. Feature Selection

Presented data in Fig.1 are just raw data gathered from accelerometer in a frequency of 4Hz. Even we haven't applied any filter or other transformation yet, it is clearly visible that a decision whether a user is walking or making another action could be based on a vector magnitude *v*.

Activities lying (on a back) and sitting are very similar, because user's position is stationary and an orientation of the smart phone is the same, with *x* axis heading to the ground and thus with value around -10. It is obvious that to clearly differentiate between these two activities by only one sensor is nearly impossible.

While standing, the *y* axis is heading to the sky and therefore takes on a value around -10.

To apply an algorithm, we extracted feature values:

- vector magnitude
- arithmetic mean sequence
- variance sequence
- minimum and maximum of mean and variance sequence

C. Heuristic Method

We define a sequence of mean values for each axis x, y, z and v as (3).

$$\forall i = 1 \ldots \frac{t \cdot f}{k} : \mu_{a_i} = \frac{1}{k} \sum_{j=ki}^{j+k} a_j \tag{3}$$

where:

- t is the number of seconds each measurement has taken, in our case $30s$
- f is the frequency of sensor's firings, in our case $4Hz$
- k is the window for mean calculation
- a is the axis of accelerometer vector, or vector magnitude v; $a \in (x, y, z, v)$

Similarly we define a sequence of variances for each axis x, y, z and v as (4).

$$\forall i = 1 \ldots \frac{t \cdot f}{k} : \sigma_i = \frac{1}{k} \sum_{j=ki}^{j+k} (a_j - \mu_i) \tag{4}$$

Next we define minimum and maximum of mean sequence (5)

$$minM_a = \min_i (\mu_{a_i}), maxM_a = \max_i (\mu_{a_i}), \tag{5}$$

and minimum and maximum of variance sequence (6)

$$minV_a = \min_i (\sigma_{a_i}), maxV_a = \max_i (\sigma_{a_i}), \tag{6}$$

To find a threshold between two characteristics we define a formula (7)

$$\tau = s_1 + \frac{s_1}{s_2} \times \mid s_1 - s_2 \mid : s_1 < s_2 \tag{7}$$

where s_1 and s_2 is an average value of minimum and maximum of any characteristics, e.g. for mean sequence of y axis it is (8).

$$s = \frac{maxM_y - minM_y}{2} \tag{8}$$

D. Decision tree

Resulting decision tree is depicted on Fig.3. Leaves of the tree are recognized activities. We have added a special value for unrecognized state *"other"*. Such an activity may be detected, when for instance a user is shaking with a mobile phone. Value in each non-leaf node represents a characteristics that is used to calculate a threshold. For example an activity *"walking"* is based on a variance of vector magnitude. Threshold in the root node is therefore calculated by a formula (7) calculated from a characteristics $minV_a, maxV_a$.

Value of k we have set to 12, which is a window of 3 seconds. Experimentally we have verified other values: $4, 8, 16 \ldots$, but values less than 12 decreased recognition accuracy and values higher than 12 didn't add any benefit.

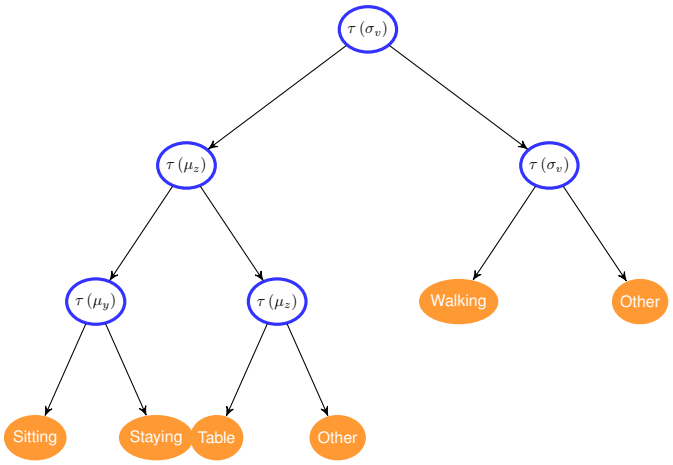


Fig. 3. Decision tree for activity recognition.

IV. Experimental Study

A. Android application

For testing purposes we have developed an application for Android operating system utilizing the aforementioned algorithm shown in Fig.4. On the main page there are only three important elements:

- currently recognized activity, e.g. *"Staying"*
- a button that starts or stops the service
- counter that logs a time duration for each particular activity

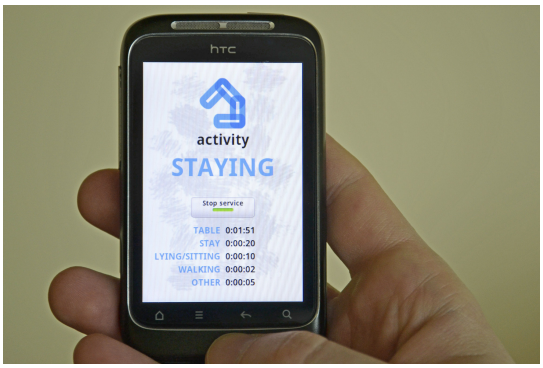


Fig. 4. Application for activity recognition for Android platform.

B. Experiments

We have conducted measurements with 25 users, where each user repeated the activity 2 times, altogether 50 measurements. The majority of users were students in the age of 22 till 28, two midle-aged users in the age of 55 and two elderly in the age of 83 and 84. The users wore a smart phone in their pocket as depicted in Fig.2 and were asked to perform each activity for 25 seconds. After performing an activity we compared time from a counter to real time the user was performing an activity. Since the user needs to push a button to start a measurement and put a phone in his/her pocket, first 5 seconds were cut off resulting in 20 seconds for each activity.

C. Results

The summary results are presented in Tab I. This table is in a form of confusion matrix where columns of the matrix

represent the instances in a predicted class and rows represent the instances in an actual class. The values are number of seconds how the real activity was recognized. For example if a user was walking for $20s$, but application recognized $18s$ of walking and $2s$ of staying, the first row would be 18, 2, 0, 0.

		Predicted class				
		Walking	Staying	Sitting	Table	Other
Actual class	Walking	**922**	67	0	0	11
	Staying	17	**936**	39	0	8
	Sitting	30	13	**952**	0	5
	Table	0	0	0	**1000**	0
	Accur.(%)	92.2	93.6	95.2	100.0	

TABLE I
CONFUSION MATRIX OF EXPERIMENTAL RESULTS

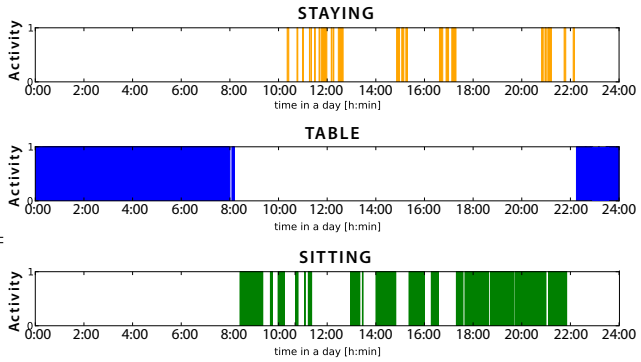


Fig. 5. Activity of a user during the day

V. CURRENT AND FUTURE WORK

We have already adopted the presented algorithm for an activity recognition by a smart phone for a long term monitoring, where the aim is to monitor following types of anomalies:

- *Unusual long inactivity* - This could happen for instance when a user doesn't wake up in the morning. Mobile phone is lying on a table when not assumed, which can indicate some health problem.
- *Unusual activity* - The example situation is, when user is sitting or walking at night, but is expected to be sleeping. It may indicate some health problem.
- *Not present when should be* - When user is expected to be at home, but still doesn't return. This is evaluated in coaction with signal strength of wi-fi.
- *Changes in daily rythm* - The user was used to have a walk in the afternoon. When just sitting in a living room without any other movement, could indicate walking or gait impairment.

Activities data are used for outlier detection service. Classification is based on three values: *activity* - activity the user is currently performing, *start time* - time when activity started, *duration* - the length of time interval, while person continues to perform the same activity (Fig. 5). To avoid noise, only activities with duration above some threshold are considered. For classification we have employed a technique *Self Organizing Maps* (SOM) for unsupervised learning, because of the fact, it does not require a number of clusters to be defined in advance. The number of activities performed during a day is unknown in advance and differs for each user.

Presented solution assumes that user would still carry a smart phone with himself, which in real conditions is hard to assure. Due to this fact, we would like to stress that presented model is more a proof of concept than an application ready for a market. The model will find its place in wristlets with accelerometers which were unveiled just now by companies like *Samsung* or *Fitbit*.

VI. CONCLUSION

In this paper we have presented an algorithm for activity recognition on a mobile device utilizing data from accelerometer sensor. By this algorithm we are able to recognize four different states: a user is walking, staying, sitting or the smart phone is lying on a table. The recognition model is based on a decision tree technique were the threshold values were calculated by heuristic approach.

Experimental study was realized by an application on HTC and Samsung smart phone with 25 users. Since the resulting service is oriented more on elderly, we currently lack more elderly users (only two users above 80 tested the application). However, we got good results also for their significantly slower gait velocity.

Currently we are using activity recognition for a long term human behavior monitoring for the purpose of anomaly detection in their daily habits, which was outlined in the last section.

ACKNOWLEDGMENT

This work is the result of the Project implementation: Competency Centre for Knowledge technologies applied in Innovation of Production Systems in Industry and Services, ITMS: 26220220155, supported by the Research & Development Operational Programme funded by the ERDF.

REFERENCES

[1] A. Zaidi, A. Sidorenko, W. D. Association, and U. S. Gallen, *Features and Challenges of Population Ageing Using the European Perspective*, ser. The WDA-HSG discussion paper series on demographic issues. World Demographic Association, 2008.

[2] M. Novak, M. Binas, and F. Jakab, "Unobtrusive anomaly detection in presence of elderly in a smart-home environment," in *ELEKTRO, 2012*, may 2012, pp. 341–344.

[3] L. Bao and S. S. Intille, "Activity Recognition from User-Annotated Acceleration Data Pervasive Computing," *Pervasive Computing*, vol. 3001, pp. 1–17, 2004.

[4] E. Tapia, S. Intille, W. Haskell, K. Larson, J. Wright, A. King, and R. Friedman, "Real-time recognition of physical activities and their intensities using wireless accelerometers and a heart rate monitor," in *Wearable Computers, 2007 11th IEEE International Symposium on*, Oct. 2007, pp. 37–40.

[5] B. Kaluža and M. Gams, "An approach to analysis of daily living dynamics," in *World Congress on Engineering and Computer Science 2010*, Oct. 2010, pp. 485–490.

[6] J. Yang, "Toward physical activity diary: motion recognition using simple acceleration features with mobile phones," in *Proceedings of the 1st international workshop on Interactive multimedia for consumer electronics*, ser. IMCE '09, New York, NY, USA, 2009, pp. 1–10.

[7] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, "Activity recognition using cell phone accelerometers," *SIGKDD Explor. Newsl.*, vol. 12, no. 2, pp. 74–82, Mar. 2011.

[8] "Fitbit: Product page," 2013, accessed: March 2013. [Online]. Available: http://www.fitbit.com/flex

[9] "Android developer page: Motion sensors," 2013, accessed: March 2013. [Online]. Available: http://developer.android.com/guide/topics/sensors/sensors_motion.html#sensors-motion-accel

Independent Hybrid Honeypot Using Passive Fingerprinting Method for Enhancing Computer System Security

[1]*Peter FANFARA (3rd year),* [2]*Ján RADUŠOVSKÝ (2nd year)*
Supervisor: [3]*Liberios VOKOROKOS*

[1,2,3] Dept. of Computer and Informatics, FEI TU of Košice, Slovak Republic

[1]peter.fanfara@tuke.sk, [2]jan.radusovsky@tuke.sk, [3]liberios.vokorokos@tuke.sk

***Abstract*—Significant amount of computers is attacked every day, either attackers alone or through automated systems. In many cases computer systems contain valuable data. Their very safety is one of the important areas in field of Information Technology (IT). Honeypot technology in combination with Intrusion Detection System (IDS) represents a solution for rapid increase of the security level against attackers from outside and inside. The proposed hybrid Honeypot is capable to adapt the environment of deployment. Right after deploying in system and using passive fingerprinting method Honeypot can create security decoy-based elements. These elements have an ability to successfully record and slow down any enemy attack or intrusion. This property is directly contributed to improving security in computer systems.**

***Keywords*—Computer System Security, Honeypot, Intrusion Detection System, Passive Fingerprinting Method.**

I. Introduction

People are able to find information and send messages quickly and easily due to the rapid spread of Internet and Web technologies. However, if we don't put a sufficiently high priority on the basic system security at the same time, hackers can take over computers using the malicious code through the existing system vulnerabilities and program weaknesses. A major damage to the most of companies and personal properties will be caused after the attacker's invasion, destruction, theft and falsification of information. Nowadays, resulting from the potential threats, a growing interest in improving the information security arises as well as the intrusion detection.

The beginnings of intrusion detection have brought some complications. There still exists a gap between the theoretical and practical level of intrusion detection. Well established defense of network/system is based on using firewall and intrusion detection system (IDS). Once the attackers are aware that the firewall has allowed an exception for the external security service they are able to use this service to gain access to the internal servers through the firewall. Subsequently, it can result in performing another attack. IDS cannot provide additional information about detection of enemy attacks and cannot reduce losses caused by those attacks [1].

A conventional approach to the security is considerably focused on defense, but the interest is increasingly devoted to more aggressive defense forms against the potential attackers and intruders. The protection against intrusions based on the bait by using a Honeypot is an example of this form [2].

Honeypot is an advanced decoy based technology which simulates weak points of system security and unsecured system services. The potential attackers focus on system vulnerabilities and very often attack the system weakest points which are simulated by Honeypot. This feature represents the nature of system security. Some of Honeypot solutions like Honeyd or Honeynets are already used in order to increase the system security [3].

The proposed client-server architecture uses a specific hybrid Honeypot that mainly consists of existing tools like Dionaea, Sebek and Snort, for rapidly increasing security in distributed computer systems. The proposed Honeypot has an autonomous feature which enables to use it in a random deployment environment. This Honeypot will auto-configure itself on the basis of system parameters obtained via passive fingerprinting method.

The following chapters describe system security using IDS with the detection mechanism based on the advanced Honeypot technology.

II. Intrusion Detection System

IDS can be defined as a tool or software application that monitors activities of computer system and/or network due to the potential occurrence of malicious activities or breaching security policy. IDS produces reports for the control station. It is primarily focused on identifying and recording information about any events as well as reporting similar attempts [4][5].

A. Classification of Intrusion Detection System

In view of the various environment applications, the IDS can be classified into two general types [1]:

- *Host-based* – consists of an agent located on host computer which is used for continuous monitoring of information from the system audit data or network activities logs. This IDS sensor type typically includes a software agent. If

there are unusual circumstances, the system automatically generates and sends a warning.

- *Network-based* – the independent platform for intrusions identification using direct capturing of transmitted network packets and monitoring several computers. Detection sensors are placed in network bottlenecks for capturing all network traffic and analyzing an individual packet contents looking out for dangerous operations.

On the basis of the detection method IDS can be divided into three types below [1]:

- *Anomaly detection* – refers to the pattern found in the data set that is inconsistent with normal behavior. The anomaly detection provides basic performance for normal network traffic. An alarm sounds only if the current network traffic is beyond/below standard parameters.
- *Misuse detection* – collects previous hacker attack characteristics and patterns which are then saved to knowledge attack database. Consequently, it can identify attacks with the same patterns and characteristics as those of previously stored attacks. IDS cannot trigger alarm if the hacker uses new attack method that hasn't been previously reported or detected.
- *Hybrid mode detection* – represents attack detection using a combination of previous two types, resulting in the reduction of generating false alarms.

B. IDS Structure and Architecture

IDS consists of several elements illustrated in Fig. 1 where the main element is a sensor – the mechanism for analysis, responsible for intrusion detection. This sensor contains a mechanism that makes decisions regarding breach. Sensor receives data from three main sources of information: IDS knowledge database, system logs and audit trails. System logs may include, e.g. file system configuration and user permissions. This information forms the basis for further decision on intrusion detection.

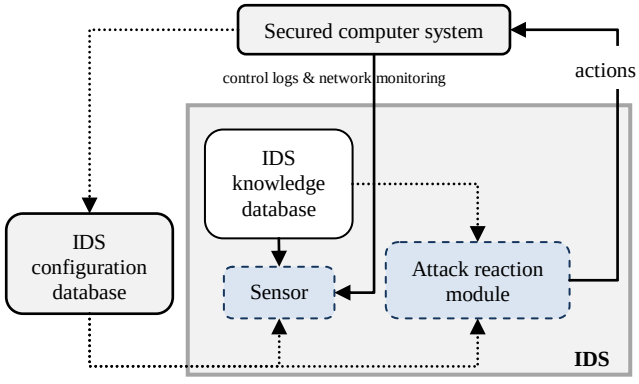


Fig. 1 Intrusion detection system structure [6]

Sensor in IDS elements illustrated in Fig. 2 is integrated together with the component which is responsible for data collection – the events generator. The data collecting method is set by policy of events generator that defines filtering method for events information notifications. The events generator (operating system, network & application) in accordance with security policies produces sets of events (system logs, control records or network packets). These occurrences may be stored together with information policy either in a protected system or outside it. In some cases they're not stored, e.g. when events streams are directly transmitted to analyzer, especially network packets [7][8].

The role of sensor is filtering information and discarding any irrelevant data obtained from event file related to protected system and detecting suspicious activity. For this purpose sensor uses the detection policy database which is composed of following parts: pattern attack, normal behavior, profiles and necessary parameters. The database contains IDS configuration parameters and communicating method with reaction module. Sensor has the custom database that also includes a dynamic history of potential intrusions [7].

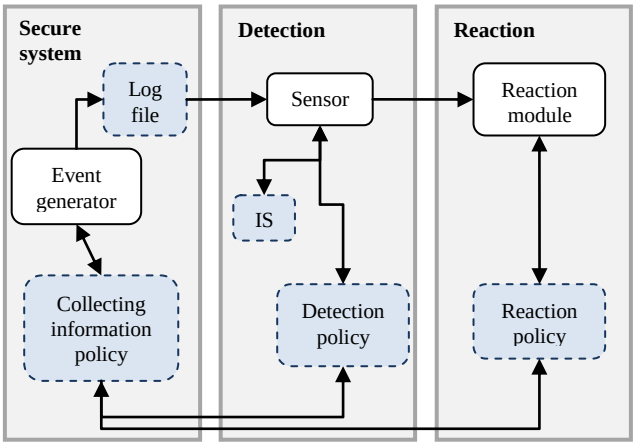


Fig. 2 Intrusion detection system elements [1]

C. Intrusion Detection Tools

Nowadays many IDS exist (i.e. Snort [9], SAX2,...), and all are specific to the system of deployment. A Snort is mostly used tool which has excellent additional conditions for usage to enhance the distributed system security in combination with Honeypots.

Snort represents the open-source IDS that can detect and warn of attack (e.g. against Honeypot). It can also capture packets and network load given by packets included in the attack. Collected information may be critical for analyzing the attacker's activities. Snort uses a modular architecture and rules based language. It combines the abnormal behavior, detection signature and different protocol detection methods [10].

Methodology of lying and cheating by providing emulation of some system services was domesticated in order to be able successfully monitor hackers activities in distributed computer systems. At first sight this system appears to be legitimate. It is possible to record and monitor all hackers' activities due to penetration and clarification of various attackers' tactics. This idea is assumed by using an advanced security tool called Honeypot.

III. HONEYPOT

Honeypot is a closely monitored network decoy available in different shapes and sizes, serving various purposes. It can be placed in a computer network with the firewall – in front of it or/and behind it. These points of deployment are the most frequent attacker's sites for obtaining access to the system. These sites provide the best solution to acquire the maximum information about attackers' activities. The main aim of Honeypot is to get information by compromising the system data in the way that any system infiltration would be

unfeasible to do in the future.

The main benefit of Honeypot is in detection. It can address IDS shortcomings – minimizing amount of generated false positive and false negative alerts. There are several situations in which IDS cannot generate a warning of attack: attack is too short, appropriate security rule refers too many false alarms or detects excessive network traffic and thus drops packets. One solution is to use Honeypot, since it has no way to affect system functions. Honeypot implementation uses unused IP address which means that all incoming communication is almost certainly unauthorized – i.e. no false positive or false negative alarm warnings or large data files to be analyzed [6].

Data obtained from the Honeypots can be used to create better protection and countermeasures or system reconfiguration against future threats.

A. Types of Honeypots

Honeypots can be classified in different ways. The classification according to purpose and level of interaction is the most frequent one.

Purpose Honeypots

This basic classification divides Honeypots based on the area of deployment.

- *Research Honeypot* – this type is used merely for research. The main objective is to obtain as much information as possible about intruder in a way that allows fully infiltration and penetration of security system. It is used to obtain information and detect new methods and types of tools used to attack other systems as well as analyzing hacker's traces, their identity or modus operandi. Another option in research; Honeypot can be used in discovering potential risks and information vulnerabilities in enterprise system [11].

  The primary function is to examine how attackers proceed and lead their attacks, which usually means understanding their motives, behavior and organization. Research Honeypots are complex in terms of deployment, maintenance and capturing huge amounts of data. On the other hand they are highly useful security tools in the field of development and in enhancing forensic analysis capabilities.

  Despite the information obtained from research, Honeypot can be used to improve prevention against attack. By improving detection and response to attack this Honeypot type totally contributes to direct security only by a small amount [12].
- *Production Honeypot* – it is used in organizations for protection and helping to reduce level of risk. It provides immediate enhancing system security [3]. Since it doesn't require such functionality as research Honeypot, its development and deployment is usually much easier. Although it can identify various attack methods. Production Honeypot provides less information about attacker than the research one. It is possible to determine where the attackers come from and what specific action was performed but it cannot determine intruders' identities, how they are organized or which tools were used.

  Production Honeypot has minimum value as a prevention mechanism. The best way to implement Honeypot is to use well-firewalled system, IDS, mechanisms for locking and fixing the system [12].

Level of Interaction

All Honeypots are based on the same concept – nobody should interact with Honeypot. The level of interaction can be defined as a maximum range of options available to attack allowed by Honeypot. Therefore, any transactions or interactions based on definition become illegitimate. Honeypots can be also categorized according to the level of interaction between intruders and the system. This classification helps in choosing the correct type for deploying in system [12].

- *Low-interaction* – doesn't contain any operating system (OS) for communication with attacker. All tools are installed purely for emulation of OS and services that cannot be used to gain full access to Honeypot. Emulation is set up to cooperate with the attacker and malicious code resulting in radical risk reduction. Attackers can only scan Honeypot and connect to several ports. Low-interaction Honeypots are characterized by possibility of easy deployment and maintenance. Honeyd is an example of low-interaction Honeypot.
- *Medium-interaction* – this type is more sophisticated than previous one but still doesn't have installed any OS. Medium-interaction Honeypot only provides an illusion of real OS to the attacker because it contains a number of emulated services the attacker can interact with. This type is able to detect automated attacks and extract information about getting malware binaries. Malicious software can be automatically downloaded and analyzed. The Dionaea tool and Honeytrap are the examples of this Honeypot type.
- *High-interaction* – the most advanced Honeypot. On the other hand, it represents the most complex and time-consuming design with the highest rate of risk, because it implies the functional OS. It gives attacker the ability to communicate with real OS where nothing is simulated, emulated or restricted. This Honeypot allows collecting the highest amount of information as it can detect and analyze all performed activities. Main accent is set to obtain valuable information about intruders by making available the entire system or even allow handling with it.

B. Architecture of Hybrid Honeypot

Hybrid Honeypot represents a combination of two Honeypots with different levels of interaction. The combination is secure solution because it is possible to take advantage of both Honeypot types that complement each other and thus limit their disadvantages shown in TABLE I.

The ideal solution is using a low-interaction Honeypot with high-interaction one. Low-interaction Honeypot acts as a lightweight proxy which relieves high-interaction Honeypot and allows focusing on processing all IP address space network traffic [3].

It is impossible for each proposed Honeypot not to use the implementation tools which have considerable importance in improving system security.

TABLE I
THE ESSENCE OF HYBRID HONEYPOT

Low-interaction Honeypot	High-interaction Honeypot	Hybrid Honeypot
+ fast	- slow	+ fast
- no possibility to detect unknown attack	+ possibility to detect unknown attack + 0 false produced warnings	+ possibility to detect unknown attack + 0 false produced warnings
+ resist to time-bomb + handles interaction with attackers	- unable to resist time-bomb and can't handle interaction with attackers	+ resist to time-bomb + handles interaction with attackers
+ cheap	- expensive	+ relatively expensive
+simple to set up and maintain	- complicated to set up and maintain	- complicated to set up and maintain

Dionaea is a modular architecture using low-interaction Honeypot. It is able to simulate server's main services and vulnerabilities due to attracting attacker/attack attention or withdrawal the malicious code [13].

Sebek is the most advanced tool for comprehensive data collection, aiming to capture as much information about attackers' activities as possible from Honeypot by stopping specific system calls (*syscalls*) on the kernel level [16].

C. Advantages and Disadvantages

All security technologies have some risk margin. If knowledge and experiences represent the power for the attackers, they also provide advantages for security professionals. By knowing the Honeypot risks it is possible to use knowledge to mitigate them and reduce disadvantages [11].

Honeypots have several unique advantages which are special for this advanced technology [3]:

- Small data sets – Honeypots can monitor only the traffic that comes directly to them. They collect small amounts of data but on the other hand, they may contain high value information.
- Minimal resources – Honeypot requires minimum system resources for capturing harmful activities. System with low-end specifications will be enough to run a Honeypot.
- Discovery new tools & tactics – Honeypots capture everything that starts the interactions with them.
- Encryption or IPv6 – Honeypot can also operate in encrypted or IPv6 environments/systems.
- Simplicity – Honeypots are very easy and flexible to operate, so they do not need complicated algorithms to function properly.

The decoy-based technology, like other security solutions, also has its own disadvantages which are described bellow [3]:

- Risk of takeover – if an attacker takes control over Honeypot, he can exploit it to attack other systems inside or outside the system of deployment.
- Limited vision – Honeypots can only monitor the traffic that comes directly to them.
- Discovery and fingerprinting – Honeypot has some expected characteristics or behavior. If attacker uses some fingerprinting tool, he can identify working Honeypot in attacked system. Even simple error, such as misspelled word in emulated service, can act as Honeypot signature.

IV. IDS ARCHITECTURE USING SOPHISTICATED HYBRID HONEYPOT

The main IDS weakness lies in the ability to detect new attack types. Use of different attack strategies or new tools could not be detected by IDS. These new attacks need to be registered in IDS configuration database and then it is possible to detect them. The proposed IDS uses hybrid Honeypot with autonomous ability to reduce the risk of detection failure and provides extensive data collecting. Hybrid Honeypot also affords the opportunity to design safety features of distributed systems through the captured data. It also minimizes any system intrusion threats. Hybrid Honeypot combines several tools: Snort, Dionaea and Sebek. The proposed system, illustrated in Fig. 3, analyzes all captured various data formats due to rapid response to attacks. It also serves as a warning reporting system to the system administrator via web interface when interaction with Honeypot occurs.

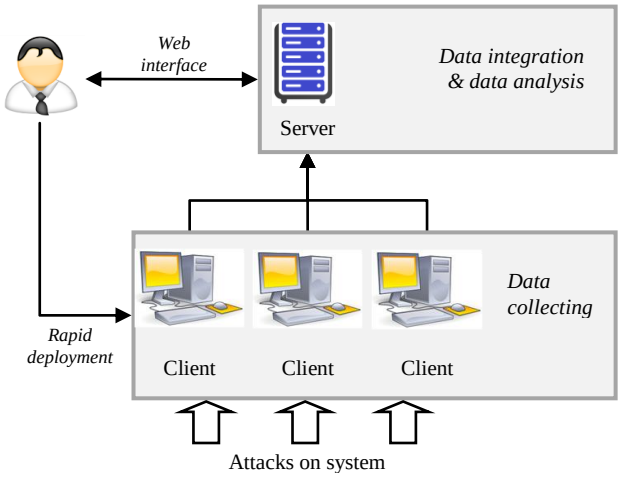


Fig. 3 Architecture of proposed detection system

The proposed intrusion detection system contains existing client-server detection architecture and its arrangements for using the proposed sophisticated Honeypot technology.

The architecture consists of several clients and a central server. Clients collect information about an attack and the captured malware is send back to server. The server records and analyzes received data, issues warning and displays overall information to the system administrator via web interface. Architecture is designed to achieve effect of centralized distributed information management and to build complex distributed system of early warning for distributed computer systems.

A. Client Architecture

The clients are installed in the same domain because of gathering data activities during the attack. Diverse system components for collecting data sets are activated depending on different type of cyber-warfare activities. Then the data sets are sent to server for further analysis and subsequently they update the system security. Client architecture shown in Fig. 4 consists of three components:

- *Snort* – monitors and filters packets during intrusion detection. It identifies patterns & characteristics of specific attacks, information and warning messages.

- *Dionaea client* – simulates general services and vulnerabilities that attract the attackers. It captures malware patterns and characters.
- *Sebek client* – records the attacker behavior during interaction with Honeypot into the log files.

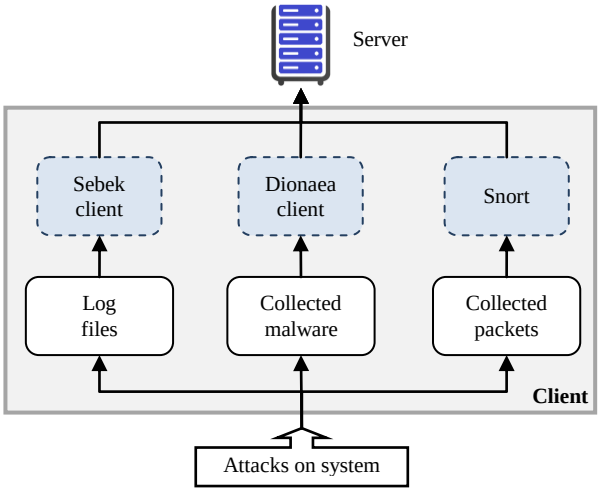


Fig. 4 Client architecture

B. Server Architecture

At the same time server is connected to multiple clients owing to centralization of collected data and it is set to receive all outgoing messages, which are then stored in the database. Server architecture is shown in Fig. 5. It indicates the attacker intention targeted to extensive computer or scanning attacks by using individual interconnection reports. Architecture of the server consists of three main parts, outputs which are normalized before they are going to be stored into the database:

- *Dionaea server* – receives malware patterns sent by Dionaea client component.
- *Sebek server* – simultaneously receives and filters multiple data sources representing instruction or cohesion of data sent to store.
- *Verification* – modular design of open-source hybrid system for detecting an intrusion using standard communication format. Verification part can receive the data from many clients and integrate disparate data formats.

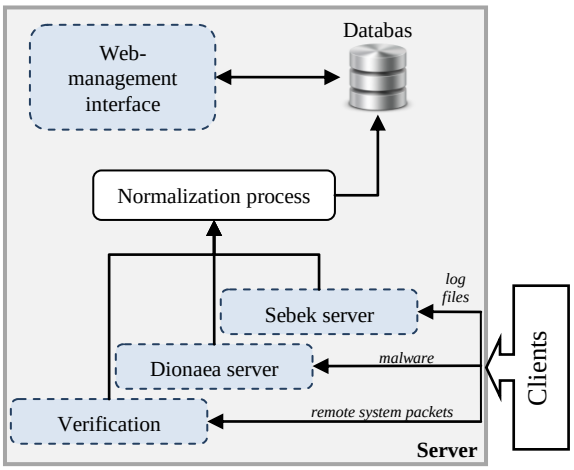


Fig. 5 Server architecture

Web server interface displays all attack analyses obtained from the database. At the same time it monitors the attack patterns and occurrences of unusual circumstances. In the event of their occurrence the specific messages are highlighted via web management interface for correct and in time response.

C. Sophisticated Hybrid Honeypot

Desired State

Sophisticated hybrid Honeypot is a suggestion addressing the current state of the security system which works on the plug-&-play principle. The optimum state will occur when Honeypot will run a complete configuration process right after plug-in. For example, after installing the Linux OS to the distributed computer system, we will have Linux Honeypot, or after removing any of services the related service will be removed from the Honeypot list of emulated services. When replacing routers in the system, for example routers from Hewlett-Packard vendor to Cisco routers, the existing Honeypot, pretending to be a router, will immediately and autonomously auto-reconfigure and update itself. The solution is a device that just simply connects to the network/system and learns the topology autonomously without any external support. Upon completion of the scanning process the device will accurately determine the number of Honeypots with their configuration and it will be able to adapt quickly to any modifications in the system.

Trouble

The most critical component of sophisticated Honeypot is the method how the Honeypot gets the information about the deployment network. E.g. what systems are used and how they are used in the current environment. There is shown the example of heterogeneous distributed computer system in Fig. 6. Honeypot will be able to sophistically map and promptly respond to the current system environment after obtaining network parameters. One of the simplest possible ways is an active probing and thus determining the system and type of used services. The use of active method of data mining also has some shortcomings in terms of increased network load – there is a risk to the running system functionality.

Sophisticated Honeypot would have to constantly scan all active environment of deployment to remedy the described lack. This solution is not the most appropriate approach.

Solution

The solution to the drawbacks of active system scanning is a passive approach, specifically passive fingerprinting and mapping method. Passive fingerprinting method is not new. The idea is to get the system overview via mapping the current environment. The difference against the active method has a different mapping approach. This approach is based on obtaining information through passively capturing network traffic, analyzing it and then determining the system identity based on unique system fingerprints. The passive method uses the same method as an active one but in different ways. Tools, such as Nmap [14], create a signature database that contains known operating systems and services. All searching tools actively broadcast packets that require a response from destination devices right after creating a signature database. Incoming responses are unique to the most operating systems and services. Responses are simply compared with the data in signature database due to a clear identification of the operating system and used services.

Passive fingerprinting uses the same approach as signature database, unless the data are obtained passively. Instead of active probing the system passive fingerprinting method intercepts network traffic and analyzes the captured packets which are then compared with a signature database. After the analyzing process ends the concrete operating system should be known. Passive fingerprinting is not limited to use only TCP protocol which allows to use the other protocols. The usage of passive method represents several advantages like: less likelihood of damage or shut down the system or service, ability to identify systems using firewall. The passive method is continuous that means changes in the network structure captures in a real time [8]. This advantage becomes a critical feature in maintaining a realistic Honeypot over a long time period. The only disadvantage of passive method represents the correctness of functioning through the routed networks – the most effective usage of passive obtaining parameters method is in local area networks.

Concept

The proposed Honeypot data obtaining mechanism is based on the concept of passive fingerprinting method. Honeypot is deployed as an independent device that is physically connected to the computer network of a distributed computer system. The tracking and learning phase starts after connecting to a network device. In this phase Honeypot learns the topology and plans the deployment of other Honeypots. Duration of learning phase is variable and depends on system topology. Proposed Honeypot can determine the number of used systems, types of operating systems and running services via passive analyzing of the network traffic. It also has the feature to determine with whom and how often a concrete system or service communicates. This information is used for mapping and obtaining knowledge about the deployment network. Once Honeypot collects all the necessary information it can start with the Honeypot deployment illustrated in Fig. 7. The created Honeypots are designed to mirror the real system and decrease the risk of the attack. Honeypots with the ability to look and behave in the same way as the production environment can easily blend with their surroundings. Their identification and tracking by attackers is much more difficult.

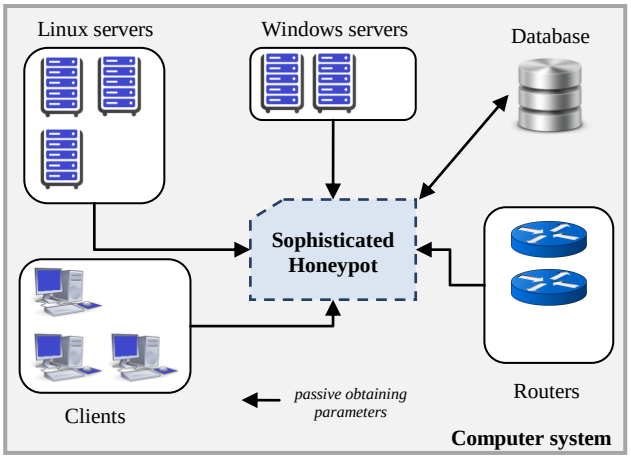


Fig. 6 Passive obtaining system parameters via sophisticated Honeypot during determination process of virtual Honeypots deployment

Passive acquisition of information does not end but it is continuous. It monitors the entire network system and increases its flexibility. Any change is identified in real-time and necessary steps (the system deployed Honeypots) are realized in the fastest possible time.

The proposed sophisticated Honeypot considerably reduces the work caused by configuration and administration in constantly changing environment.

Deploying Honeypots in a System

The traditional solution to the issue of implementing Honeypot to the system requires the physical placement of a new computer for each monitored IP address. Physical Honeypot deployment represents considerable period of time and work. Autonomous, simpler solution e.g. fire-&-forget, is not to implement a physical Honeypot but virtual type which, if in sufficient quantity, can monitor all unused IP addresses. Virtual decoys pursue identical IP address space as the system itself. All virtual decoys are designed, located and managed by only one physical device – the proposed sophisticated Honeypot illustrated in Fig. 7.

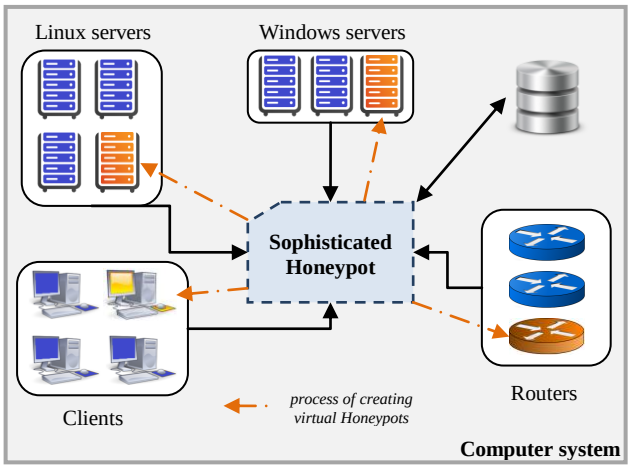


Fig. 7 Deployment of virtual Honeypots based on obtained parameters

Whereas virtual Honeypots monitor unused IP addresses in computer networks, it is almost certain that any activity detected on the monitored IP addresses is most likely a malicious or unauthorized behavior. Using information gathered through passive mapping of the environment can determine the quantity and type of Honeypot deployment.

The ability to dynamically create and deploy virtual decoys already exists. Open-source solution with a low interaction Honeypot called Honeyd [15], allow deploying virtual environments decoys throughout the organization. It is possible to realize the design of sophisticated autonomous Honeypot with dynamic creating and deployment of virtual decoys that minimizes the risk of detection and identification of intruders by merging surrounding environment with combination of options such a Honeyd solutions capabilities and passive fingerprinting.

V. Conclusion

Security of information technologies is essential in the society that depends on information. Therefore, considerable emphasis is placed on data and information sources protection in systems developing processes. Protection of access, availability and data integrity represents the basic safety features required for information resources. Any disruption of these properties will end in penetration into the system

and would increase security risk. One way of the defense is a system that detects unusual and suspicious behavior called IDS. IDS major risk is represented by undetected penetration problem.

An advanced technology called Honeypot has huge potential for the security community and it can also achieve several objectives of other technologies that make it almost universal solution. Usage of Honeypots represents a cost-effective answer to improvements in the organization security status.

Honeypots, like any new technology, also have some shortcomings that need to be overcome and removed. Use of the decoy based technology represents a cost-effective solution to increase the security status of the organization. Therefore they are being deployed in systems in increasing rate, but mostly like a passive device. Many system administrators monitor the situation in the system via Honeypot and if a production environment was attacked once, administrators analyze and implement solutions manually – Honeypots capabilities are not used at all or they are used minimal. Despite of many advantages of Honeypot it is not a panacea for breaching system security. Since it is used for gathering information about attacker and other threats, it is useful as an IDS detection mechanism.

The future of Honeypots and cyber security intrusion detection lies in sophisticated (autonomous hybrid) decoys. They have a radical revolutionary assumption in autonomous deployment and maintenance. They are becoming a highly-scalable solution due to their network studying and monitoring real-time feature. Deployment and management becomes more cost-effective and also provides better integration into the system of deployment. Another advantage of proposed Honeypot lies in minimizing the risk of human errors during manual configuration. Merger surrounding environment also minimizes the risk of being identified by the attacker.

ACKNOWLEDGMENT

This work has been supported by the Slovak Research and Development Agency under the contract No. APVV-0073-07 (20%). This work is also a result of implementing the Project: Development of the Centre of Information and Communication Technologies for Knowledge Systems (ITMS project code: 26220120030) supported by the Research & Development Operational Program funded by the ERDF (80%).

REFERENCES

[1] J. McHugh, A. Christie, J. Allen: "Defending Yourself: The Role of Intrusion Detection System," IEEE Software, IEEE Computer Society, pp. 42-51, October 2000.

[2] F. G. Lyon: "Nmap Network Scanning: The Official Nmap Project Guide to Network Discovery and Security Scanning," [online], Nmap Project, USA, ISBN 978-0979958717, January 2009. Available on: <http://nmap.org/book>.

[3] S. Karthik, B. Samudrala, A. T. Yang: "Design of Network Security Projects Using Honeypots," Journal of Computing Sciences in Colleges, 2004.

[4] L. Vokorokos, N. Ádám, A. Baláž: "Application of Intrusion Detection Systems in Distributed Computer Systems and Dynamic Networks," Computer Science and Technology Research Survey, Košice, 2008, pp. 19-24, ISBN 978-80-8086-100-1.

[5] L. Vokorokos, N. Ádám, A. Baláž, J. Perháč: "High-performance Intrusion Detection System for Security Threats Identification in Computer Network," Computer Science and Technology Research Survey, Košice, 2009, pp. 54-61, ISBN 978-80-8086-131-5.

[6] R. Baumann, C. Plattner: "White Paper: Honeypots," Swiss Federal Institute of Technology, Zurich, 2002.

[7] L. Vokorokos et al: "Architecture of Intrusion Detection System Based on Partially Ordered Events", Computer Science and Technology Research Survey, Vol. 2, 2007, pp. 80-91, ISBN 9788080860714.

[8] L. Vokorokos, A. Pekár, N. Ádám: "Data Preprocessing for Efficient Evaluation of Network Traffic Parameters," INES 2012: IEEE 16th International Conference on Intelligent Engineering Systems, 2012, Lisbon, Portugal, pp. 363-367, ISBN 978-1-4673-2695-7.

[9] Snort [online]. Available on: <http://www.snort.org>.

[10] M. Tomášek, M. Čajkovský, B. Madoš: "Intrusion Detection System Based on System Behavior", SAMI 2012: 10th IEEE Jubilee International Symposium on Applied Machine Intelligence and Informatics: proceedings: Herľany, Slovakia, 2012, pp. 271-275, ISBN 978-1-4577-0195-5.

[11] L. Spitzner: "The Value of Honeypots, Part One: Definitions and Values of Honeypots," Security Focus, 2001.

[12] L. Spitzner: "Honeypots: Tracking Hackers," Boston, USA: Addison-Wesley, Pearson Education, 2003, ISBN 0-321-10895-7.

[13] Dionaea catches bug [online]. Available on: <http://dionaea.carnivore.it/>.

[14] R. Chandran, S. Pakala: "Simulating Network with Honeyd," [online], Technical paper, Paladion Networks, December 2003. Available on: <http://www.paladion.net/papers/simulating_networks_with_honeyd.pdf>.

[15] N. Provos: "Developments of the Honeyd Virtual Honeypot," [online]. Available on: <http://www.honeyd.org>.

[16] Sebek [online]. Available on: <http://www.honeynet.org/tools/sebek/>.

[17] P. Jakubčo, E. Danková: "Distributed Emulation Using GPGPU," Proceeding of the Faculty of Electrical Engineering and Informatics of the Technical University of Košice, Slovakia 2011, pp. 284-287, ISBN 978-80-553-0611-7.

[18] FANFARA, P., et al. Autonomous Hybrid Honeypot as the Future of Distributed Computer Systems Security. Unpublished.

Industrial robot optimization for required accuracy and speed

[1]*Peter PAPCUN (2nd year)*
Supervisor: [2]Ján JADLOVSKÝ

[2,3]Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

[1]peter.papcun@tuke.sk, [2]jan.jadlovsky@tuke.sk

***Abstract* — This paper describes design of industrial robot optimization for required accuracy and speed. I began with analysis in article Optimizing industrial robot for maximum speed with high accuracy at conference Modeling of mechanical and mechatronic systems. This article continues with analysis of robot movement and then paper dedicates design of mentioned optimization. Control and optimization of industrial robot's speed and accuracy is very important nowadays, for example in quality welding or application of sealants and many other examples. I am working with industrial robot MELFA RV-2SDB made by Mitsubishi.**

***Keywords* — Industrial robot, speed optimizing, accuracy optimizing, required speed, required accuracy**

I. Introduction

This document analyzes robot movement and design robot optimization for required speed and accuracy based on mentioned analysis. Article [1] started with analysis of industrial robot movement.

This article is divided to seven chapters. The first chapter is introduction certainly. The second chapter describes design of movement point-to-point and present types of movements. Next chapter dedicates trajectories and their parametric expression. The fourth chapter describes results of movement point-to-point. The fifth chapter describes optimization methods for required speed and accuracy. The sixth chapter analyzes results of movement point-to-point and describes design of optimizing. The last chapter is conclusion.

II. Design of movement point-to-point

Paper [1] dedicates analysis of industrial robot movements on linear and circular trajectories. These movements use intern functions of robot controller. Article [1] describes data collections from industrial robot for movement analysis. I use this data collection again for movement point-to-point. I remind types of movements from paper [1], because I use same types of movements in method point-to-point:

- **SX** - linear movement in X-axis direction (coordinates Y and Z do not change during the motion)
- **SY** - linear movement in Y-axis direction (coordinates X and Z do not change during the motion)
- **SZ** - linear movement in Z-axis direction (coordinates X and Y do not change during the motion)
- **SA** - linear movement which all coordinates (X, Y, Z) change
- **CX** - circular movement in plane which is parallel with plane YZ
- **CY** - circular movement in plane which is parallel with plane XZ
- **CZ** - circular movement in plane which is parallel with plane XY
- **CA** - circular movement in the whole space (all coordinates is changing during motion)

Data analysis use mentioned tools in paper [1]: application for analyze robot accuracy and application for analyze robot speed. Movement speed does not be limited with any way. Point-to-point movement will be programmed with method, that every point will be exactly defined which robot's effector (endpoint of industrial robot) has to pass. These points will be calculated through parametric equation. Distance between points will be depended on step of iteration. I describe parametric equations of trajectories in the third chapter by this equations will be calculated points of trajectories.

III. Kinematics and trajectories

This chapter describes some equations of trajectories. I write expressions of trajectories in plane with functions and parametric equations. Trajectories are written through parametric equations in space. Parameter t can be time or step of iteration in parametric equations.

Line:

Plane: Function: $y = kx + q$ (1)

k – slope of a line

q – line shift

Parametric equation: $x = a + ct$

$y = b + dt$ (2)

[a,b] – point on line

(c,d) – direction vector of line

Space: $x = a + dt$

$y = b + et$

$z = c + ft$ (3)

[a,b,c] – point on line

(d,e,f) – direction vector of line

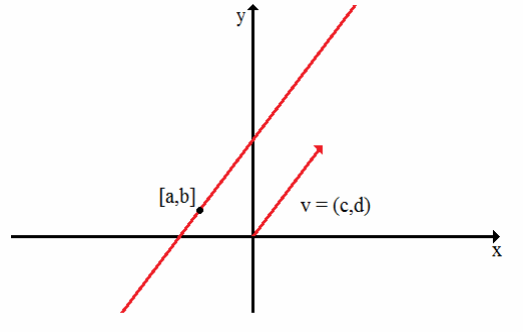


Fig. 1: Line on plane

Circle on plane:

Equation: $(x-a)^2 + (y-b)^2 = r^2$ (4)

Parametric equation: $x = a + r\cos t$

$y = b + r\sin t$ (5)

[a,b] – centre of circle

r – circle radius

Ellipse on plane:

Equation (canonical form):

$$\frac{(x-x_0)^2}{a^2} + \frac{(y-y_0)^2}{b^2} = 1 \quad (6)$$

Parametric equation: $x = x_0 + a\cos t$

$y = y_0 + b\sin t$ (7)

$[x_0, y_0]$ – centre of ellipse

a, b – length of semi-axis

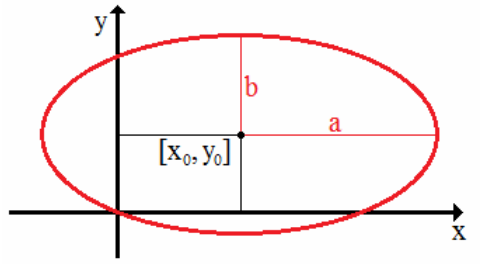


Fig. 2: Ellipse on plane

Circle and ellipse on space:

Parametric equation: $x = x_0 + a\cos t$

$y = y_0 + b\sin t$

$z = z_0 + c\cos t$ (8)

$[x_0, y_0, z_0]$ – centre of ellipse or circle

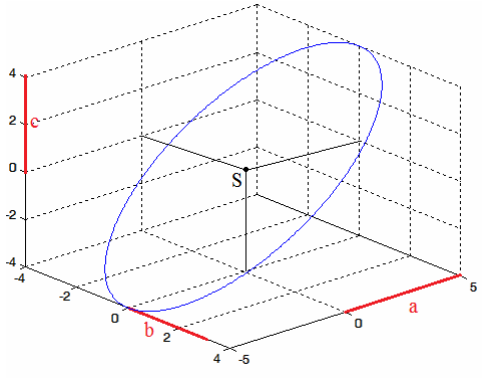


Fig. 3: Ellipse on space

Spiral:

Parametric equation: $x = x_0 + t\cos t$

$y = y_0 + t\sin t$ (9)

Helix:

Parametric equation: $x = x_0 + r\cos t$

$y = y_0 + r\sin t$

$z = kt$ (10)

k – twist

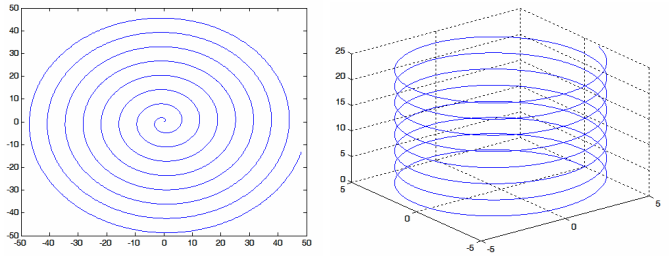


Fig. 4: Spiral on plane and helix on space

Trajectories are many others, but in this part are only some. Many others trajectories can be mathematical expressed, through parametric equation.

IV. Movement point-to-point

I test line movement at the first. I determine specific parametric equations of lines, intervals and steps of iteration (t). Simple program (in program language C) calculates points of movement from defined specific parametric equation (3). Step of iterations are determined gradually so that distance between points are 0.01 mm, 1 mm, 5 mm, 10 mm, 25 mm, 5 cm, 1 dm.

If distance was 0.01 mm, than industrial robot was maximally accurate (maximum accuracy is 0.01 mm – sensors restriction), but speed was very low, lower then 1 mm/s. When distance was 1 mm, than industrial robot was maximally accurate and average speed was 61 mm/s, these speeds was in interval from 47.04 mm/s to 71.23 mm/s. If distance was 10 mm, robot accuracy was 0.1 mm and average speed 231 mm/s. The highest distance was 1 dm. When distance was 1 dm, then speeds was higher than was measured in [1], but accuracy was very low (2 mm and near of system centre 2 cm). We can see a summary of results in the following table:

Table 1: Line movement which use method point-to-point

Typ. of mov.	Distance [mm]	Max. deflect. [mm]	Min. speed [mm/s]	Max. speed [mm/s]	Average speed [mm/s]	Measured speed [mm/s]
SX	1	0.00	60.10	64.23	63.57	771.56
	10	0.06	230.53	241.42	237.21	
	100	0.98	761.47	863.15	821.32	
SY	1	0.00	60.61	63.22	62.17	639.54
	10	0.05	233.40	242.46	236.51	
	100	1.02	645.41	701.12	674.74	
SZ	1	0.00	66.51	71.23	68.27	832.18
	10	0.08	203.80	213.07	209.45	
	100	2.45	830.47	915.12	875.24	
SA	1	0.00	47.04	51.71	49.21	882.45
	10	0.10	235.20	259.38	242.32	
	100	18	895.12	952.13	924.21	

Notes to tables 1 and 2:

Typ. of mov. – type of movement

Distance – distance between the nearest points of trajectory

Max. deflect. – the highest measured movement error (deflection)
Min. speed – the highest measured speed
Max. speed – the lowest measured speed
Average speed – average speed of all measured speeds
Measured spd. – the highest average speed measured in [1] (movements which use intern functions of robot controller)

Then I test circular movement. Program (in program language C) calculates points of circular movements, through parametric equations of circular trajectories (7, 8). Iteration steps are determined as in the previous case so that distance between points of trajectory are 1 mm, 5 mm, 10 mm, 25 mm, 5 cm, 1 dm. Measurement results:

Table 2: Circular movement which use method point-to-point

Typ. of Mov.	Distance [mm]	Max. deflect. [mm]	Min. speed [mm/s]	Max. speed [mm/s]	Average speed [mm/s]	Measured speed [mm/s]
CX	1	0.00	61.66	68.84	65.14	606.47
	10	0.09	225.78	236.21	232.29	
	100	8.1	690.68	701.64	696.12	
CY	1	0.00	60.31	68.21	64.61	529.5
	10	0.08	226.82	235.16	231.45	
	100	9.4	681.18	692.38	687.23	
CZ	1	0.00	57.47	65.22	61.11	554.21
	10	0.15	209.24	225.42	216.47	
	100	12.48	573.99	583.85	579.35	
CA	1	0.00	51.57	59.68	56.16	599.51
	10	0.27	226.79	243.53	234.27	
	100	24.17	643.23	672.34	654.58	

We can see in this table, that accuracy with distance 10 mm is worse than in previous case (linear movement) and accuracy with distance 1 dm is not accurate (accuracy 24.17 mm). If distance was 1 dm, average speed was better than was measured in [1], this is same result as in previous case. When we compare rows in tables 1 and 2 with same distance, we can see similar average speeds. This implies that speed depends on distance between points of trajectory regardless of movement type and this distance depends on iteration step.

V. Optimization

Effector speed will be optimized through density of defined points on required trajectory. These points will calculate online in program loop. This loop will be programmed in robot controller. Trajectory will be defined by parametric equation in mentioned loop. New point of trajectory will calculate in each loop. Previous chapter describe iteration step, this step will modify so that robot will have required speed and accuracy. User will be able to define either of parametric equation in function or text file, also equation which is not mentioned in the third chapter.

Industrial robot optimizes for required speed and accuracy will operate by three ways:

1. Optimization for required speed.
2. Maximum speed optimization for required accuracy.
3. Speed optimization for required accuracy and speed.

Optimization for required speed has defined speed as input. Program modify iteration step to value, that effector has required speed regardless of movement accuracy. The first iteration step modifies according point-to-point movement analysis, which we can see in next chapter. If speed does not be equal with defined speed after this modify, than program start with fine tuning of iteration step.

Maximum speed optimization for required accuracy has defined accuracy as input. Program modify iteration step to value, that effector has required accuracy. The first iteration step modifies according point-to-point movement analysis, too. If accuracy does not be equal or better with defined accuracy after mentioned modify, than program reduce iteration step. If accuracy is equal or better with defined accuracy, than program increase iteration step minimally, until effector exceed accuracy. If program get twice same value of iteration step, then robot has maximum speed with required accuracy.

Speed optimization for required speed and accuracy has defined speed and accuracy as input. This optimization works same as optimization for required speed. If robot have required speed after optimization for required speed, then program check accuracy. If accuracy is equal or better with defined accuracy, then optimizing is at the end. When accuracy does not be equal or better with defined accuracy, program informs about this fact. When we would like required accuracy, we can continue with maximum speed optimization for required accuracy.

VI. Analysis of Movement Point-to-Point

We can see graph on figure 5 where are relationship between trajectory points distance and effector speed. Effector speeds are average measured values from movement point-to-point (average values from table 1 and 2).

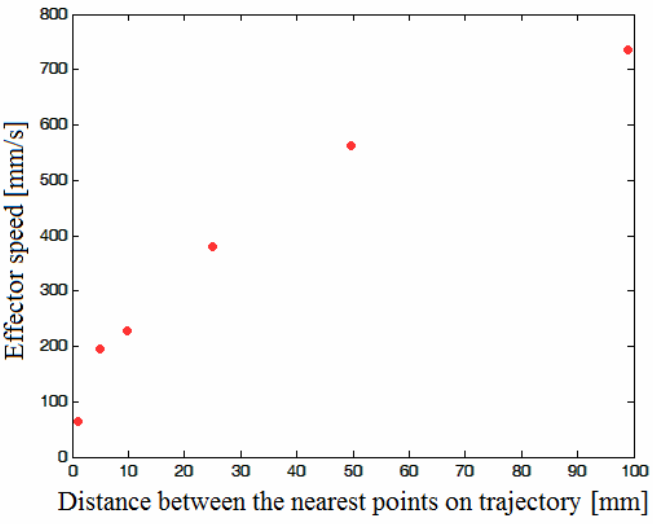


Fig. 5: Relationship between effector speed and trajectory points distance

Values which are in graph (Fig. 5):

Table 3: Values in graph, relationship between speed and distance

Distance [mm]	1	5	10	25	50	100
Speed [mm/s]	61.3	193.1	230.0	380.5	565.6	739.1

I approximate points on graph by method of least squares. I mark effector speed as v and distance between the nearest points on trajectory as s. Result is this function:

$$v = - 0.0657s^2 + 13.0036s + 92.6894 \quad (11)$$

We can see approximated function (11) on figure 6:

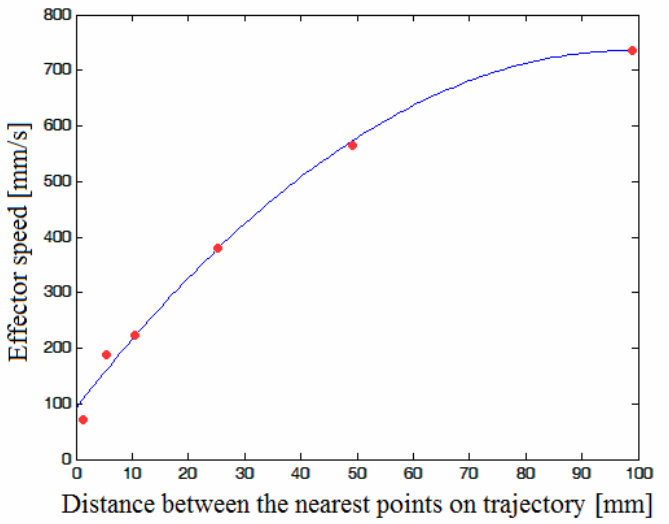


Fig. 6: Relationship between effector speed and trajectory points distance with approximated function

Now, we can define required speed and calculate distance s approximately, by approximated function (11).

I can do similar operation with accuracy. We can see average values of maximum deflection (the highest measured movement error) as accuracy from tables 1 and 2 in table 4:

Table 4: Relationship between trajectory points distance and accuracy

Distance [mm]	1	5	10	25	50	100
Accuracy [mm]	0.00	0.03	0.11	0.59	2.37	9.58

Approximated function to values in table 4 by method of least squares (p – accuracy, s – distance):

$$p = 0.001s^2 + 0.001s + 0.01 \quad (12)$$

Points from table 4 and approximated function (12):

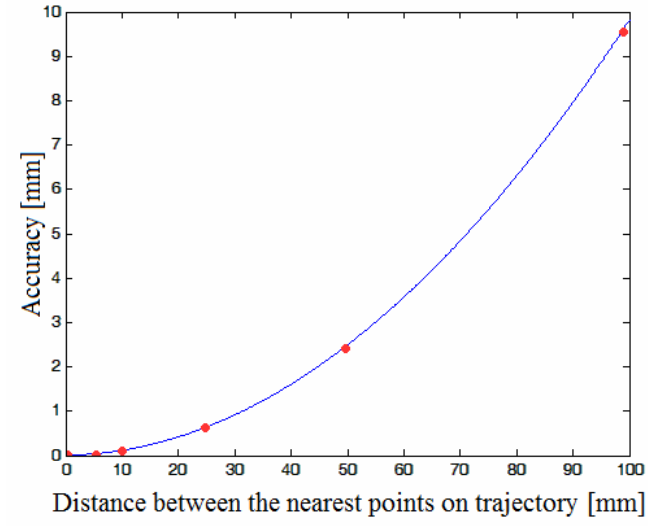


Fig. 7: Relationship between accuracy and trajectory points distance with approximated function

We can calculate distance s approximately from required accuracy by approximated function (12).

I described three method of optimization in previous chapter. I mentioned in each of them, that initial state of iteration step (the first iteration step modify) program calculate from analysis movement point-to-point. Optimization for required speed and speed optimization for required accuracy and speed use function (11) and maximum speed optimization for required accuracy use function (12). Program calculates distance between the nearest points on trajectory mentioned as s from functions (11) and (12). Trajectory is specified by parametric equation, so I can use formula (13) for calculate length of curve.

$$l = \int_0^{t_{max}} \sqrt{x'(t) + y'(t) + z'(t)}dt \quad (13)$$

When I know length of curve l and distance s, than I can calculate iteration step k (t_{max} is maximum value of parameter t from parametric equation of curve):

$$k = t_{max} : \frac{l}{s} = \frac{t_{max} s}{l} \quad (14)$$

Now, I have calculated iteration step, this step will use first at optimizing effector speed and accuracy as initial state of iteration step. Next, program start tuning of iteration step, according procedure described in chapter five. Concept fine tuning mean modify iteration step about value t_{max}/l.

VII. Conclusion

I designed method of speed and accuracy optimization in this paper. At the first I analyzed robot movement by method point-to-point. Then I designed optimization. When I was designing optimization, I suggested concept initial state of iteration step. I can calculate initial step based on the measured values from analysis of movement point-to-point. This calculation is described in the sixth chapter. My research will continue with realization of this optimization design, also I would like design and realize diagnostic tool which will control speed and accuracy.

Acknowledgment

This work has been supported by the Scientific Grant Agency of Slovak Republic under project Vega No.1/0286/11 Dynamic Hybrid Architectures of the Multiagent Network Control Systems.

References

[1] JADLOVSKÝ, J. – PAPCUN, P. 2012. Optimizing Industry Robot for Maximum Speed with High Accuracy. In: Procedia Engineering. No. 48 (2012), p. 533-542. - ISSN 1877-7058

[2] PAPCUN, P. – ČOPÍK, M. – ILKOVIČ, J. 2012. Riadenie robota integrovaného v pružnom výrobnom systéme. In: ElectroScope. Vol. 2012, no. 2 (2012), p. 1-9. - ISSN 1802-4564

[3] MOSTÝN, Vladimír – KRYS, Václav: Mechatronika průmyslových robotů, Ostrava, 2012, ISBN 978-80-248-2610-3

[4] BOŽEK, Pavol – BARBORÁK, Oto – NAŠČÁK, Ľubomír – ŠTOLLMANN, Vladimír: Špecializované robotické systémy, Bratislava, 2011, ISBN 978-80-904766-8-4

[5] Instruction manual, CRnQ/CRnD Controller, Mitsunishi Electric, Ratingen, Germany, 2010.

[6] Product leaflets, RV-2SDB, Mitsubishi Electric, Ratingen, Germany, 2010.

[7] PAPCUN, P. 2011. Control of robot integrated in flexible production line, diploma thesis, Košice, Slovakia, 2011.

Jitter Utilizing Inter-flow Communication Environment – JUICE

[1]*Ivan KLIMEK (3rd year),* [2]*Marek ČAJKOVSKÝ (2nd year)*
Supervisor: [3]*Assoc. Professor František JAKAB, PhD*

Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

[1]ivan.klimek@tuke.sk, [2]marek.cajkovsky@tuke.sk, [3]frantisek.jakab@tuke.sk

***Abstract*—In this paper we show that TCP flow synchronization, resp. congestion packet loss synchronization in general, despite the current status quo considering it a negative consequence of the TCP/queuing design in fact demonstrates and opens completely new possibilities of congestion control. We present a proof-of-concept first-ever protocol enabling inter flow communication without infrastructure support thru a side channel constructed on generic FIFO queue behavior. This enables independent flows passing thru the same bottleneck queue to collaborate and achieve fair capacity sharing and a stable equilibrium state.**

***Keywords*— congestion control, transport protocol, inter-flow communication, jitter, flow synchronization**

I. Introduction

TCP flow synchronization occurs when congestion caused packet drop reaches levels that it affects all the flows on the overflowing bottleneck, as TCP understands packet drop as an indication of congestion it reduces its data rate. Because all the affected flows react the same way, the effect is flow synchronization. [1,2] More generally speaking this behavior is not TCP-specific, congestion loss can be utilized for synchronizing into a stable equilibrium state, where instead of aggressively reducing the rate as TCP does, flows stop increasing their rate and only reduce it so that packet loss doesn't occur, thus stabilize the overflowing buffer, reaching and holding the full capacity of the bottleneck. Protocol that utilizes this behavior was already constructed [3], the problem with this state is that the capacity is not distributed fairly amongst flows – it is just stabilized at whatever utilization the flows reached in the moment bottleneck capacity was reached. Research into jitter-based congestion/rate control shows that any change in the available capacity will be visible on the change of jitter.[3] While experimenting with this stable state we noticed an interesting phenomena, we were able to on demand break the stable state and increase or decrease jitter levels on all flows passing thru the bottleneck with the change of the data rate of one of the competing flows even if it represented a fraction of the total bottleneck data rate. Thus effectively creating a broadcast communication medium between independent flows from the bottleneck queue.

II. Changing the Perspective

State of the art transport protocols still use the TCP "point of view" - looking at data on packet level [4,5], even if Forward Error Correction (FEC) protocols slowly gain traction [6] they are no different. The presented design is based on FEC codes and uses their error correcting properties to change the perspective from a packet level to flow level. This abstraction can be compared to the difference in the physics of a particle and a wave. We no longer have to deal with individual packets, retransmit them etc. they are all just pieces of a flow. Where the flow can be modulated as a waveform to create a secondary channel without affecting the primary channel – the information it transmits.

The goal we set to achieve with the possibility of communication between flows is to enable fast convergence to a stable network state where resources are allocated with absolute fairness. This is in strong contrast with currently used TCP-based logic, where flows compete for capacity in a never ending fight for resources without ever reaching a stable state or full resource utilization. All designs that tried to improve this situation did so by utilizing feedback from the network requiring the network to change – routers to add functionality. [7,8] The presented system is the first as far as authors are aware of that provides the capability for network flows to cooperate in achieving optimal network utilization without the need for any support from the network.

Most queues still use the most basic queuing algorithm – FIFO/DropTail [9]. This fact that FIFO queues are "everywhere" in Internet is considered an issue causing problems such as Bufferbloat [10]. In fact their omnipresence is a crucial enabler for our mechanism to function. Other queuing mechanisms such as AQM/RED/WRED etc should be also theoretically to some extent supported but we did not so far look deeper into them.

The presented design primarily targets last mile routers as they are where most congestion occurs [10]. This scope limitation enables us to design a simple proof of concept communication system as not too many concurrent flows occur at a last mile bottleneck when compared to a Internet backbone router bottleneck situation. Figure 1 illustrates the core concept behind our design, independent flows interact at the bottleneck queue. When the ingress packet rate is higher than egress packet rate – buffering occurs. Packets are ordered inside of the queue by their ingress time, if one flow increases its rate, the gap between packets of other flows will be altered creating a measurable change. In the previously mentioned

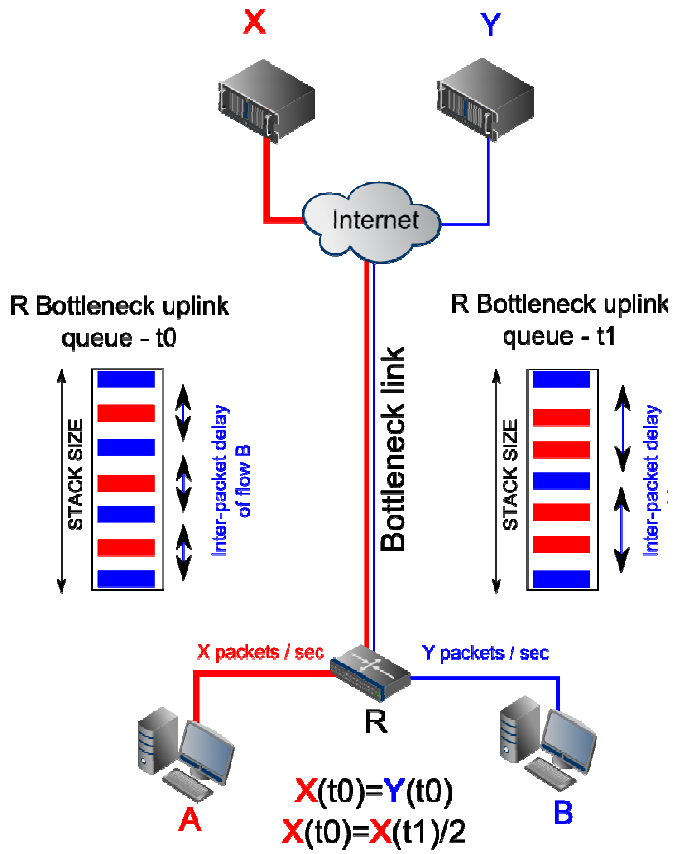


Fig. 1. FIFO inter-flow interaction

equilibrium state achieved by JUICE [3] a change of the TX inter-packet delta aka packet rate of one flow can on-demand break the queue equilibrium and create jitter measurable by all flows in that queue. The flow can also return back to the original rate, thus a stable/unstable jitter modulation scheme can be constructed.

III. FIFO THE THEORETICAL COMMUNICATION CHANNEL

FIFO is far from a perfect communication medium, it is not linear as the TX of a frequency or amplitude modulation will not necessarily create a linear response on the RX of a different flow. The sampling rate of the RX flow is not stable and not only a function of its packet rate but a function of the current state of the queue that is a superposition of all the different flows passing thru it. Classic modulation approaches such as frequency or amplitude modulation are thus unusable, transmitting special patterns such as patterns with strong auto-correlation properties also doesn't produce stable results as error, erasures and shifts are arbitrary. Even if frequency modulation would be possible it would not be ideal as it would require many samples to be precise. We therefore propose a modulation requiring only a few samples that provides deterministically achievable results – an equilibrium breaking modulation – where flows reach a stable state that is broken and restored in protocol defined time windows.

Because of the mentioned properties a robust coding scheme is required, first of all the true sampling rate is not known therefore the code needs to be self-clocking. The start of the sequence needs to be clearly recognizable, to achieve this the presented code is self-synchronizing by reversing the transmitted sequence every period.

$$xwy \in X^* \Rightarrow xw, wy \in X^* .$$

Equation. 1. Self-synchronization, a code X over an alphabet A has a synchronizing word w in $A+$

Self-clocking is achieved by utilizing Differential Manchester encoding where a transition is guaranteed at least once every bit, allowing the receiver to perform clock recovery. Differential scheme is preferred as detecting transitions is often less error-prone than comparing against a threshold in a noisy environment as a jitter based modulation is, also the polarity of the transition is unimportant just the presence of a change is detected. The receiver monitors the variance of jitter, where jitter is defined as the inter-packet delta. Variance is derivative of jitter, thus the second derivative of packet arrival time.

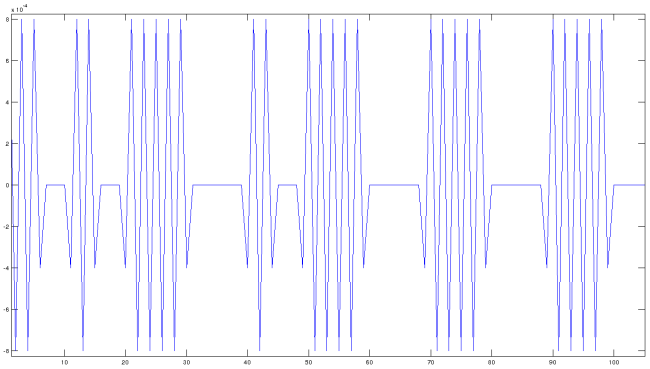


Fig. 2. Sample delta jitter of a receiver flow

The variance is measured by taking the zero-crossings of the delta jitter. The number of zero crossing directly reflects the state of the queue – stable/not stable. Because the modulation window is protocol defined, the number of variance changes per window can be interpreted to bits.

The presented design uses 2 changes to encode bit 1, one change to encode bit 0 – the reconstructed signal in Figure 3 is therefore 11001000.

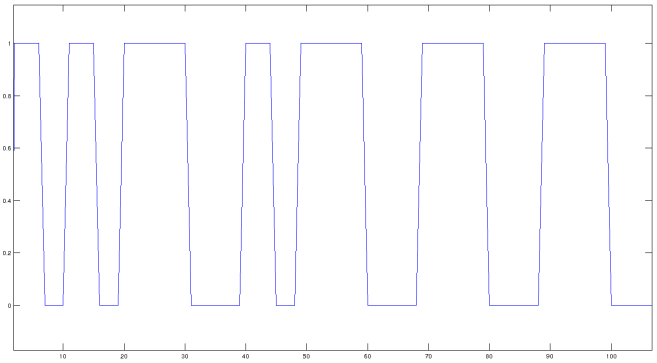


Fig. 3. Sample receiver reconstructed signal - Diff. Manchester encoded

The simulations were performed in ns2 using a star topology with five TX nodes, one router node and one receiver node, each TX node was transmitting one flow. All flows merged at the router node which connection with the receiver node represented the bottleneck link. The modulated flow changed its inter-packet delta between 4ms for the stable state and 2ms for the excited state. The four receiver flows had a constant inter packet delta of 2ms. Protocol defined

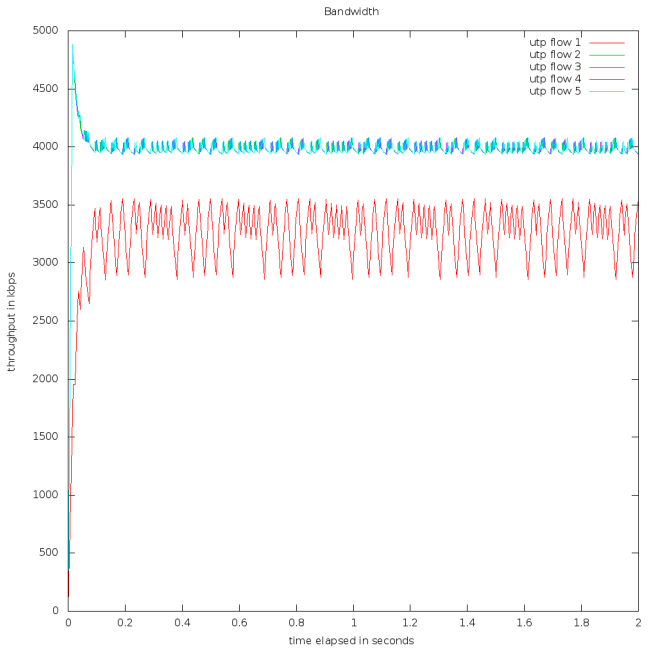


Fig. 4. Bottleneck bandwidth utilization per-flow - red is the modulated flow, blue are the 4 receiver flows superimposed

modulation window was 20ms. Bottleneck link speed was set to match the stable state. Receiver code post-processed the ns2 data in Matlab.

Figure 4 illustrates the bandwidth allocation of the bottleneck per flow, the 4 receiver flows are superimposed as they have the same rate, so the total rate of the not-modulated flows is packet every 0.5ms vs. packet every 2 resp 4ms for the modulated flow. Meaning in average there is 6 times more not-modulated packets than modulated in the queue at every moment.

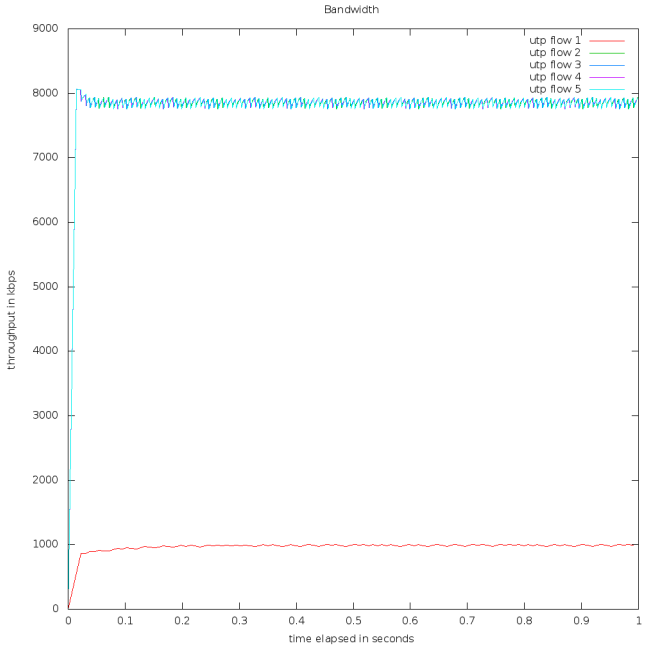


Fig. 5. Bottleneck bandwidth utilization per-flow - red is the modulated flow, blue are the 4 receiver flows superimposed

To demonstrate that only a fraction of the total queue capacity needs to be modulated for other flows passing thru it

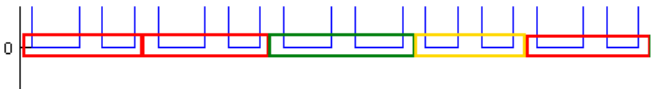


Fig 6 Reconstructed zoomed color coded signal - red is 1, green and yellow are 0 in up/down position resp.

be able to receive the signal the following experiment was performed. The not-modulated flows transmitted a packet every millisecond, the modulated flow every 8ms with oscillation of 1ms up/down resulting in 7ms/9ms inter-packet delta for duration of protocol defined window 20ms. Thus in average there was 32 times as many not-modulated packets than modulated in the queue at every moment. Figure 5 illustrates this bandwidth allocation and Figure 6 shows the reconstructed signal.

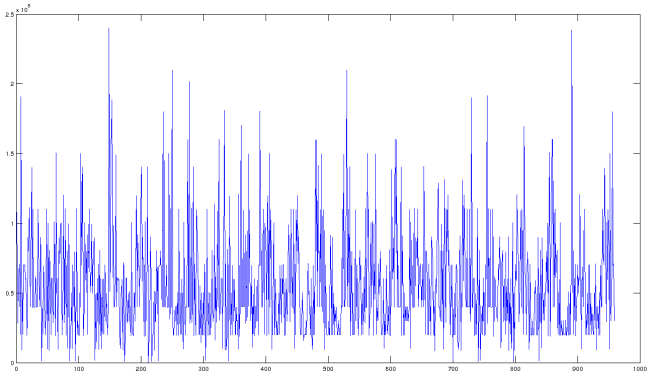


Fig 7 Jitter of one of the host B flows. Periodic modulation is clearly visible

Because the presented modulation only aims to trigger instability in the queue, only a small change is necessary to achieve measurable difference. In contrast to usual modulations where signal is being modulated on a medium, we modulate by creating and disturbing a stable state of the medium. This approach could be compared to the resonant frequencies known from physics, where even a small periodic driving force can produce large amplitude oscillations.

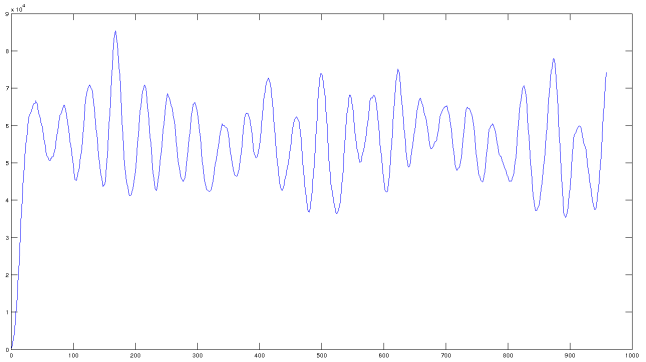


Fig 8 Filtered signal from the jitter on Figure 7

To set the amplitude right we propose to use an Automatic Gain Control on the RX of the modulated flow, because its the RX of the flow that is modulating it can have the knowledge of the fact that modulation should be happening and what should be transmitted. If the signal is not strong enough because of ambient jitter the RX can immediately notify the TX that amplitude needs to be increased. For the primary information channel the modulation/amplitude itself is invisible as it just oscillates to both direction so in average the

rate is constant.

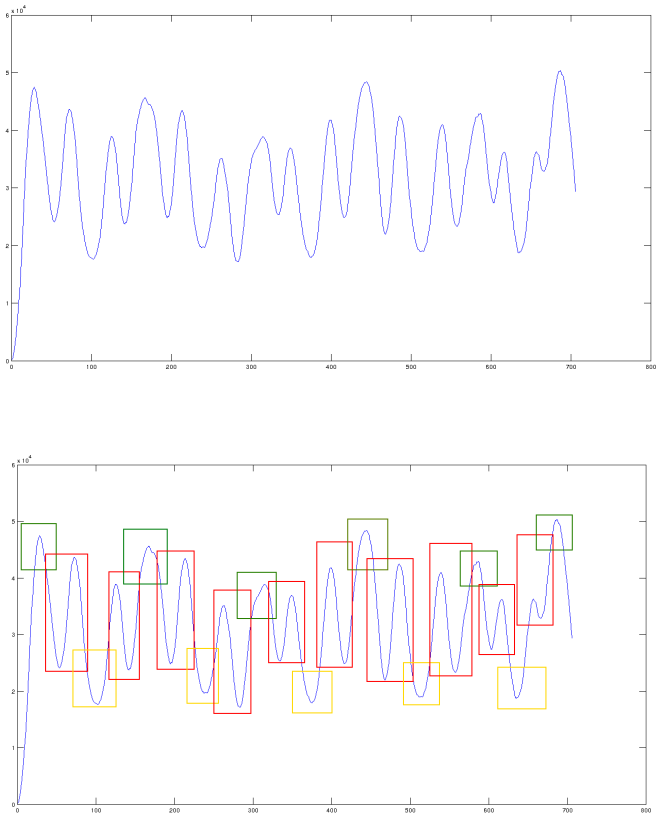


Fig 9 Signal + colored representation - bit 1, two changes per period - red, bit 0 up position - green, bit 0 down position - yellow

IV. FIFO THE PRACTICAL COMMUNICATION CHANNEL

To verify that the presented theoretical model really works, we implemented it in C and tested "in the wild" on the most worst case scenario we could think of. Worst case meaning a scenario where arbitrary jitter is abundant. We used the topology from Figure 1. Two hosts connected via WiFi to a 3G router, one server was located at the Technical University Košice, the other in Germany – approximately 20ms RTT from the server at our University. Host B had 3 active connections, each at 30ms inter-packet delta, all not-modulated – meaning static rate. Host A had a single modulated flow at base rate of 40ms, with symmetrical amplitude of 10ms, thus one period inter-packet delta was 30ms, the next 50ms and vice versa. The flows from host B and host A met at the shared bottleneck – 3G wifi router uplink. Except of that they had nothing in common, they were destined for different servers and originated from different hosts. Figure 7 shows the jitter measured on one of host B flows. Figure 8 the signal filtered from the jitter using signal averaging (two pass moving average with window size of half of the protocol defined window).

The performed experiment clearly demonstrated that only a fraction of the total traffic passing thru the bottleneck needs to be modulated for other flows to be able to detect it. In this experiment only the clocking signal was transmitted e.g. all zeros – one change per period. Figure 9 illustrate how does a 1010 signal look – two changes per period marking bit 1.

V. CONCLUSION

This paper demonstrated both using simulations and real world measurements that inter-flow communication without any support from the infrastructure can be achieved. The presented approach does not claim to be optimal, it is meant to be purely a proof of concept. Our primary goal is to provoke further discussion and gain visibility for this new idea in hope that the signal processing aspects of our approach will be improved by experts on that subject.

ACKNOWLEDGMENT

This work is the result of the Project implementation: Competency Centre for Knowledge technologies applied in Innovation of Production Systems in Industry and Services, ITMS: 26220220155, supported by the Research & Development Operational Programme funded by the ERDF.

REFERENCES

[1] L. Zhang, S. Shenker, and D.D. Clark, Obervations on the Dynamics of a Congestion Control Algorithm: The Effects of Two Way Traffic, ACM SIGCOMM '91, Zurich, 1991.

[2] C. So-In, "Loss Synchronization of TCP Connections at a Shared Bottleneck Link," WUSTL Technical Report, 2006.

[3] I. Klimek, M. Keltika: Jitter Utilizing Congestion Estimation (JUICE), Proceedings of Scientific Conference of Young Researchers (SCYR) 2012, FEI TU Kosice, 2012.

[4] G. Yunhong, R. L. Grossman. "End-to-End Congestion Control for High Performance Data Transfer.", IEEE/ACM Transaction on Networking, 2003.

[5] J. Hyungsoo, S. Kim, H. Y. Yeom, S. Kang, and L. Libman. "Adaptive delay-based congestion control for high bandwidth-delay product networks." In INFOCOM, 2011 Proceedings IEEE, pp. 2885-2893. IEEE, 2011.

[6] K. MinJi, M. Médard, and J. Barros. "Modeling network coded TCP throughput: A simple model and its validation." In Proceedings of the 5th International ICST Conference on Performance Evaluation Methodologies and Tools, pp. 131-140. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2011.

[7] D. Nandita, N. McKeown, and A. G. Fraser. "RCP-AC: Congestion control to make flows complete quickly in any environment." In INFOCOM 2006. 25th IEEE International Conference on Computer Communications. Proceedings, pp. 1-5. IEEE, 2006.

[8] Q. I. Ayyub, T. Znati, and L. Andrew. "Congestion control using efficient explicit feedback." In INFOCOM 2009, IEEE, pp. 10-18. IEEE, 2009.

[9] R. Narendran, and C. Mala. "A Comparative Study of Different Queuing Models Used in Network Routers for Congestion Avoidance." Information Technology and Mobile Communication (2011): 431-434.

[10] J. Gettys, and K. Nichols. "Bufferbloat: Dark buffers in the internet." IEEE Queue 9, no. 11 (2011): 40.

Linear term calculus in informatics

[1]*Emília DEMETEROVÁ (1^{st} year),* [2]*Veronika SZABÓOVÁ (1^{st} year)*

[1,2]Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

[1]emilia.demeterova@tuke.sk, [2]veronika.szaboova@tuke.sk

***Abstract*—This work gives a view of the problematic related to linear term calculus. The main part of this work is about introduction of types and terms into intuitionistic linear logic. Classifying category is used as a typing system of the language. Its objects are linear types and its morphinism are linear terms. For expressing the semantics of classifying category is used symmetric monoidal closed category. There are also included some examples describing how to use the deduction rules for prove the formulae.**

***Keywords*—Linear logic, linear term calculus, type theory, symmetric monoidal closed category**

I. Introduction

Nowadays are exists several types of logics, which are different in many properties. The main differences are in their expressing, in the manner of their work with resources, that are manipulated in control manner. This field of computer science is still evolving. The logician are continually searching for new logics that satisfied the actual needs of the computer science. One of the younger logic is linear logic introduced by Girard in 1987 [7]. This logic is based on the main idea that the sources are limited. As well as in the real life, if some source is used it is no more available. This is the main reason of the fact that this logic is used for modelling real life situations.

In the last twenty years the linear logic was being evolved permanently and new logics were infered from it. One of this derived logic is intuitionistic linear logic (ILL). This logic has the same properties as the linear logic except two connectives - intuitionistic linear logic does not contain the multiplicative disjunction and the negation. ILL can be represented in the sequent calculus as a sequent. This logic allows only one formula derivable from the assumption. This means that only one conclusion is available. This is the main reason, why is not allowed to use multiplicative disjunction which needs more conclusions.

Linear logic nor intuitionistic logic does not contain types. Types are important in the checking of the computer program. Types can be used to check if the program is typed and working correctly. Without types we can control the program only during the compilation. Types into λ-calculus (which was originally untyped) was introduced by Alonzo Church [11]. This λ-calculus we can now call typed λ-calculus.

Linear term calculus is used mainly for the programming languages, which are not resource-oriented. It was introduced by a collective of authors- Hyland, Bierman, Benton and de Paiva [3]. Linear term calculus can be used to determine the propertities of some programming languages. It is a system for assigning a term for ILL.

We can find the word category in several fields of our life, e.g. in mathematics, logics, informatics, etc. Categories are a suitable tool for describing the meaning of the logic.

Linear type theory (LTT) and a linear term calculus (LTC) are encapsulated to the classifying category, which objects are linear types and the morphinism between these objects are linear terms.

The classifying category can be mapped into symmetric monoidal closed category (SMCC) by appropriate functor. A SMCC is a suitable abstract frame for assigning a meaning for object and morphinism.

II. The language of intuitionistic linear logic

The formal syntax of ILL consists of symbols, BNF grammar and deduction system. A linear formula φ is the form written by the following BNF grammar:

$$\varphi ::= p \mid 1 \mid 0 \mid \bot \mid \top \mid \varphi \otimes \psi \mid \varphi \& \psi \mid \varphi \oplus \psi \mid \varphi \multimap \psi \mid \varphi^{\bot} \mid !\varphi \mid ?\varphi$$

A symbol p denotes an atomic proposition. The formula $\varphi \otimes \psi$ denotes that both actions φ and ψ will be done. Tensor product $\otimes$ we read times , its neutral element is the constant 1. The formula $\varphi \,\&\, \psi$ describes that only one of the actions φ or ψ will be processed and we can choose which one. We can read the direct product $\&$ as with . The constant $\top$(top) is the neutral element of it. This type of connectives expresses external nondeterminism. By the formula $\varphi \oplus \psi$ we can express direct sum $\oplus$ that we read plus . Its neutral element is the constant 0. The direct sum describes that only one of the actions φ or ψ will be processed, but we cannot determine which one. The formula $\varphi \multimap \psi$ describes that an action ψ is caused by φ where φ is considered as a resource exhausted by linear implication and no more available after performing it. ! is a modal operator, which is also called of course. $!\varphi$ means unlimited utilisation of the resource φ. The second modal operator is ?, which we can read as why not. $?\varphi$ indicates the possibility of resource utilisation. By using these modal operators can be classical implication $\varphi \to \psi$ adopted into linear logic in this way: $(!\varphi) \multimap \psi$. In linear logic the negation $()^{\bot}$ is the only one negative connective. $\varphi^{\bot}$ means exhaustion of a resource φ.

The deduction system of linear logic is given by sequent calculus defined by Gentzen. The sequent has the form $\Gamma \vdash \varphi$, where $\Gamma = (\varphi_1, \ldots, \varphi_n)$ is the finite sequence linear formulae and φ is a linear formula derivable from assumptions Γ. Deduction rules are described e.g. in the work [7], [4].

Example 2.1: This example demonstrate the usage of deduction rules in intuitionistic linear logic. If the $!((\varphi \otimes \tau) \multimap \psi)$, $!((\psi \otimes \theta \otimes \theta) \multimap (\tau \otimes \theta))$, φ, τ, θ, $\theta \vdash \tau \otimes \theta$ is formula in the sequent calculus in the intuitionistic linear logic, the proof of this formula can be constructed as a proof

tree by using the deduction rules of linear logic.

$$\dfrac{\dfrac{\dfrac{\dfrac{}{\varphi \vdash \varphi}\,(\mathrm{id}) \quad \dfrac{}{\tau \vdash \tau}\,(\mathrm{id})}{\varphi, \tau \vdash \varphi \otimes \tau}\,(\otimes\mathrm{-r}) \quad \dfrac{\dfrac{\dfrac{}{\psi \vdash \psi}\,(\mathrm{id}) \quad \dfrac{}{\theta \vdash \theta}\,(\mathrm{id})}{\theta, \psi \vdash \psi \otimes \theta}\,(\otimes\mathrm{-r}) \quad \dfrac{}{\theta \vdash \theta}\,(\mathrm{id})}{\theta, \theta, \psi \vdash \psi \otimes \theta \otimes \theta}\,(\otimes\mathrm{-r}) \quad \dfrac{\dfrac{}{\tau \vdash \tau}\,(\mathrm{id}) \quad \dfrac{}{\theta \vdash \theta}\,(\mathrm{id})}{\tau, \theta \vdash \tau \otimes \theta}\,(\otimes\mathrm{-r})}{\tau \otimes \theta \vdash \tau \otimes \theta}\,(\otimes\mathrm{-l})}{\psi \otimes \theta \otimes \theta \multimap (\tau \otimes \theta),\ \theta,\ \theta,\ \psi \vdash \tau \otimes \theta}\,(\multimap\mathrm{-l})}{(\varphi \otimes \tau) \multimap \psi,\ (\psi \otimes \theta \otimes \theta) \multimap (\tau \otimes \theta),\ \varphi,\ \tau,\ \theta,\ \theta \vdash \tau \otimes \theta}\,(\multimap\mathrm{-l})$$

$$\dfrac{((\varphi \otimes \tau) \multimap \psi),\ !((\psi \otimes \theta \otimes \theta) \multimap (\tau \otimes \theta)),\ \varphi,\ \tau,\ \theta,\ \theta \vdash \tau \otimes \theta}{!((\varphi \otimes \tau) \multimap \psi),\ !((\psi \otimes \theta \otimes \theta) \multimap (\tau \otimes \theta)),\ \varphi,\ \tau,\ \theta,\ \theta \vdash \tau \otimes \theta}\,(\mathrm{dereliction})\,(\mathrm{dereliction})$$

III. Linear term calculus

The correspondence between calculations and proofs in intitionistic logical systems was introduced by Haskell Brooks Curry and William Alvin Howard. This correspondence [8] is known as Curry-Howard correspondence, which is also called as formulae-as-types or proof-as-programs. This correspondences is valid only for intuitionistic logical systems.

Based on Curry-Howard correspondence [10], the intuitionistic propositional logic is in straightforward correspondence with typed λ-calculus. The first order ILL is in correspondence with the Church's theory of types (theory of primitive types) and the second order logic is in the correspondence with polymorphic type theory. The main idea of the correspondence is the existence of a proof corresponding with a program that indicates the correctness of it.

A. Linear type theory

Programs are processing data. Every data have some type. This is the main reason why we must at first define LTT before introducing terms into ILL.

Untyped λ-calculus was defined by Church [11]. The main disadvantages of the untyped $\lambda-$calculus was that programs were possible to control only during compilaton. Church solved this problem and created the typed λ-calculus as an alternative for the $\lambda-$calculus. In typed λ-calculus the basic types have the same role as sets in set theory.

Types can be diveded into:

- basic types: such as integer, real, boolean;
- Church's linear types: functional and product linear type
- other types: such as polymorphic types, recursive types etc.

Church's linear types denoted by σ, τ, χ we define by the following production rule:

$$\sigma ::= 1 \mid X \mid \sigma \otimes \tau \mid [\sigma, \tau],$$

where

- 1 is unit type;
- X are all basic types $\mathcal{T}$ from the signature $\Sigma(\mathcal{T}, \mathcal{F})$ where $\mathcal{F}$ is finite set of function symbols $f : X_1, \ldots, X_n \to X_{n+1}$;
- $\sigma \otimes \tau$ is a product linear type;
- $[\sigma, \tau]$ is a function linear type, that is the set of functions from a type σ to the type τ. We denote by *ChT* the set of all linear Church's types.

Except functional type we can introduce special function on types, which is called linear combinators. Combinator is a function, which does not contain any free variables and expresses the type properties:

$$\begin{gathered} Id_\sigma : \sigma \to \sigma; \\ assl_{\sigma,\tau,\chi} : \sigma \otimes (\tau \otimes \chi) \to (\sigma \otimes \tau) \otimes \chi; \\ assr_{\sigma,\tau,\chi} : (\sigma \otimes \tau) \otimes \chi \to \sigma \otimes (\tau \otimes \chi); \\ swap_{\sigma,\tau} : \sigma \otimes \tau \to \tau \otimes \sigma; \\ open_\sigma : \sigma \to 1 \otimes \sigma; \\ close_\sigma : 1 \otimes \sigma \to \sigma; \\ eval_{[\sigma,\tau]} : [\sigma, \tau] \otimes \sigma \to \tau \end{gathered} \tag{1}$$

To define linear term calculus we assume for every type $\sigma \in ChT$ a countable infinite set $var(\sigma)$ variables of the type σ. For every linear Church's type σ let $preterm(\sigma)$ is a set of all preterms type σ defined as follows:

- $() \in preterm(1)$ is the empty preterm;
- if $x \in var(\sigma)$ is a variable, then $x \in preterm(\sigma)$ is the linear preterm of type σ;
- if $s \in preterm(\sigma)$ and $t \in preterm(\tau)$ are linear terms, then $(s, t) \in preterm(\sigma \otimes \tau)$ is a linear preterm of product linear type $\sigma \otimes \tau$;
- if $s \in preterm(\sigma)$ is linear preterm, $\alpha : [\sigma, \tau]$ is a function or combinator then $\alpha(s) \in preterm(\tau)$ is linear preterm of type τ.

A preterm $s \in preterm(\sigma)$ defined as above is a *term* of a type σ if no variable occurs *more than once* in s.

The term x is called as *basic linear term* if it contains no combinators.

Variables in linear terms can be substituted by the other linear terms. Let s be a term, $s \in term(\sigma)$ and x be a variable of the same type $x \in var(\sigma)$. We denote a term t by $t[s/x]$ where all occurences of x are replaced by s using the following rules:

$$()[s/x] = ()$$

$$y\,[s/x] \quad = \begin{cases} \mathrm{s} & \text{if} \quad x = y \\ \mathrm{y} & \text{if} \quad x \neq y \end{cases}$$

$$(t, u)[s/x] = (t[s/x], u[s/x])$$

$$\alpha(t)[s/x] = \alpha(t[s/x])$$

where t, u are terms of type and α is a function symbol or combinator between appropriate types. If terms t and s have no common variables then $t[s/x]$ is a term.

Example 3.1.1:

a) A preterm s of the form $s = (f(x), eval_{\sigma,\tau}(f, y))$ where $f : [\sigma, \tau]$, $x, y \in var(\sigma)$, $\sigma, \tau \in ChT$ is a term of the type $\tau \otimes \tau$;

b) A preterm t of the form $t = swap_{\sigma,\sigma}(x, x)$, where $t = swap_{\sigma,\sigma}(x, x)$, for all $\sigma \in ChT$ is not a term, only a preterm of type $(\sigma \otimes \sigma)$ because the variable x has two occurences. □

Example 3.1.2:

a) A pair of variables (x, y) where $x \in var(\sigma)$ a $y \in var(\tau)$ is a Church's type (variable of product type $\sigma \otimes \tau$) as follows

$$v = (x, y) \in term(\sigma \otimes \tau)$$

b) A term $s = (f(x), eval_{\sigma,\tau}(f, y))$ is not a basic term because it contains the combinator $eval$. □

To construct linear type theory we have to introduce a relation of equivalency on linear terms. The symbol $\equiv_\sigma$ introduces it. To define that terms $s, t \in term(\sigma)$ are equivalent

$$s \equiv_\sigma t$$

if and only if s and t have the same variables of the same types.

The appropriate inference rules are mentioned in [1],[3],[4], [5], [6], [9].

Let s, t, u are terms of the appropriate types and α, β, γ are combinators or function symbols. Then a set Ax of axioms consists of the following equalities:

$$\begin{aligned}
&Id_\sigma(s) \equiv s\\
&(\gamma \circ \alpha)(s) \equiv (\alpha(s))\\
&assl_{\sigma,\tau,\theta}(s, (t, u)) \equiv ((s, t), u)\\
&assr_{\sigma,\tau,\theta}((s, t), u) \equiv (s, (t, u))\\
&swap_{\sigma,\tau}(s, t) \equiv (t, s)\\
&open_\sigma(s) \equiv ((), s)\\
&close_\sigma((), s) \equiv s\\
&eval_{\sigma,\tau}(\Lambda(\alpha)(s), t) \equiv \alpha(s, t)
\end{aligned}$$

We denote by $LTC(\mathcal{T}, \mathcal{F}, Ax)$ the linear type theory over the set $\mathcal{T}$ of basic types, where $\mathcal{F}$ is a set of function symbols with combinators and Ax is a set of axioms. Then we can to construct the classifying category over the linear type theory LTC which objects are linear types and morphinism between the types are linear terms.

IV. Categorical semantics of intuitionistic linear logic

In the literature exist a lot of approaches to the semantics of ILL. The first is the phase semantics defined by Girard [7]. In this work we prefer symmetric monoidal closed category as a semantics for introducing ILL.

A *symmetric monoidal closed category* is a quintuple $(\mathcal{C}, \otimes, I, \epsilon, \delta)$ where

- $\mathcal{C}$ is a category;
- $\otimes : \mathcal{C} \otimes \mathcal{C} \to \mathcal{C}$ is a tensor functor;
- I is terminal object in $\mathcal{C}$, which is also neutral element of the tensor product;
- closedness is defined by the following property: for every object A in $\mathcal{C}$ and the functor $\_ \otimes A$ exist the hom-functor $Hom(A, \_)$ with a natural transformations:

$$\begin{aligned}
&\epsilon_{A,B} : Hom(A, B) \otimes A \to B\\
&\delta_{A,B} : A \to Hom(B, A \otimes B)
\end{aligned}$$

A. The model of the linear term calculus

We can interpret the classifying category $\mathcal{C}l(LTC)$ in a symmetric monoidal closed category as follows.

At first, we define the interpretation function

$$i : \mathcal{T} \to \mathcal{C}_{obj}$$

which assign to every basic type an object of category $\mathcal{C}_{obj}$. Then the representation of the basic types in symmetrical monoidal closed category as follows:

$$i(\sigma) = [\![\sigma]\!].$$

Next we extend this mapping to linear Church's types as follows:

$$\begin{aligned}
&i(I) = [\![I]\!]\\
&i(\sigma \otimes \tau) = [\![\sigma]\!] \otimes [\![\tau]\!]\\
&i([\sigma, \tau]) = Hom([\![\sigma]\!], [\![\tau]\!])
\end{aligned}$$

For interpretation of function symbols and combinators in the category $\mathcal{C}$ we must define the function interpretation mapping

$$j : \mathcal{F} \to \mathcal{C}_{morp}$$

which assign to every function symbol from signature a morphinism in category $\mathcal{C}$. Then for $f : \sigma_1, \ldots, \sigma_n \to \tau$ holds

$$j(f) : [\![\sigma_1]\!] \otimes \ldots \otimes [\![\sigma_n]\!] \to [\![\tau]\!]$$

We can extend this function interpretation mapping to the combinators as follows:

$$\begin{aligned}
&j(Id_\sigma) = id_{[\![\sigma]\!]}\\
&j(assr_{\sigma,\tau,\chi}) = a_{[\![\sigma]\!],[\![\tau]\!],[\![\chi]\!]}\\
&j(assl_{\sigma,\tau,\chi}) = a^{-1}_{[\![\sigma]\!],[\![\tau]\!],[\![\chi]\!]}\\
&j(swap_{\sigma,\tau}) = c_{[\![\sigma]\!],[\![\tau]\!]}\\
&j(open_\sigma) = l_{[\![\sigma]\!]}\\
&j(close_\sigma) = l^{-1}_{[\![\sigma]\!]}\\
&j(eval_{\sigma,\tau}) = \epsilon_{[\![\sigma]\!],[\![\tau]\!]}
\end{aligned}$$

The composition of combinators α and β, where $\alpha : \sigma \to \tau$ and $\beta : \tau \to \theta$ we denote as $\alpha \circ \beta$. We can interprete as composition of morphinism in category $\mathcal{C}$:

$$j(\alpha \circ \beta) = j(\alpha) \circ j(\beta)$$

We can conclude that an interpretation of linear term calculus $\mathcal{C}l(LTC)$ in a symmetric monoidal closed category $(\mathcal{C}, \otimes, I, \epsilon, \delta)$ is a pair of functions

$$(i, j) : \mathcal{C}l(LTC) \to (\mathcal{C}, \otimes, I, \epsilon, \delta)$$

if every basic linear term t with the equivalence $\alpha(t) \equiv \beta(t)$ derivable from the set Ax then the interpretation of combinators is the same morphinism $j(\alpha) \equiv j(\beta)$ in category $\mathcal{C}$.

Example 4.1.1:

For example we have linear term

$$swap(g, swap(swap(y, f), s))$$

of the type

$$eval([!f, eval(g, f)], eval(y, eval(f, y)))$$

and

$$g : swap(\varphi, \psi)$$
$$y : \varphi$$
$$f : [\varphi, \psi]$$
$$s : !\varphi$$

We define the type context as

$$\Gamma = (g : \varphi \otimes \psi, y : \varphi, f : [\varphi, \psi], s : !\varphi)$$

By the combinators in (1) we can express

$$g \otimes (y \otimes (f \otimes s))$$

as the term of type

$$((!(\varphi \multimap \psi)) \multimap (\varphi \otimes \psi))) \otimes (\varphi \otimes (\varphi \multimap \psi) \otimes \varphi)$$

There linear function type $f : [\varphi, \psi]$ we can express as $f : \varphi \multimap \psi$.

Semantic of a given term we constructed in SMCC where to every type we realize interpretation function i and morphisms we constructed by interpretation mapping j as follows: Let $i : \mathcal{T} \to C_{obj}$ then

$$i(\varphi) = [\![\varphi]\!],$$
$$i(\psi) = [\![\psi]\!],$$
$$i(\varphi \otimes \psi) = [\![\varphi]\!] \otimes [\![\psi]\!],$$
$$i(!\varphi) = Hom([\![\varphi]\!], [\![\varphi]\!]),$$
$$i([\varphi, \psi]) = Hom([\![\varphi]\!], [\![\psi]\!]).$$

Let $j(f) : [\![\sigma_1]\!] \otimes \ldots \otimes [\![\sigma_n]\!] \to [\![\tau]\!]$ then for $f : \varphi \multimap \psi$ is $f : \varphi \to \psi$ and

$$j(f) = [\![\varphi]\!] \to [\![\psi]\!],$$
$$j(swap_{\varphi,\psi}) = c_{[\![\varphi]\!],[\![\psi]\!]}$$

We construct SMCC as follows
Objects:

$[\![\varphi]\!]$ as the representation of φ,
$[\![\psi]\!]$ as the representation of ψ,
$[\![\varphi]\!] \otimes [\![\psi]\!]$, as the representation of linear product type $\varphi \otimes \psi$.

Morphinism:

on object $[\![\varphi]\!]$ the homeset from $[\![\varphi]\!]$ to $[\![\varphi]\!]$ as the representation of type $!s$,
on product object $[\![\varphi]\!] \otimes [\![\varphi]\!]$ as the representation of type $(\varphi \otimes \psi)$ together with projections ϕ_1 and ϕ_2,
the homeset from $[\![\varphi]\!]$ to $[\![\psi]\!]$ as the representation of type $[\varphi, \psi]$.

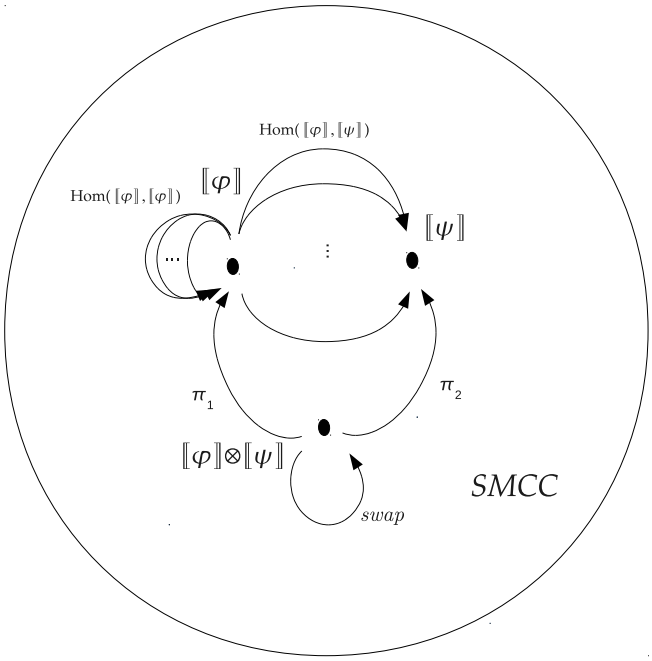


Fig. 1. Term $swap(g, swap(swap(y, f), s))$ in SMCC

V. Conclusion

The goal of our work was to describe the usage of terms in intuitionistic linear logic. This work concern about the term's origin, propertities as well as the formal description of the intuitionistic linear logic. Our plans to the future is to formulate linear predicate logic and its usage in computer science.

References

[1] J. S. Ambler, "First order linear logic in symmetric monoidal closed categories", *The University of Edinburgh*, vol. 3, pp. 15–64, 1991.
[2] S. Abramsky, "Computational interpretations of linear logic", *Technical Report 90/20, Department of Computing, Imperial College*, vol. 3, pp. 1–15, 1990.
[3] G. Bierman, N. Benton, V. De Paiva, and M. Hyland, "Term assignment for intuitionistic linear logic", *University of Cambridge*, vol. 3, pp. 15–64, 1992.
[4] G. Bierman, "On intuitionistic linear logic", *University of Cambridge*, vol. 3, pp. 15–64, 1994.
[5] N. Benton, "Strong normalisation for the linear term calculus", *Computer Laboratory, University of Cambridge*, vol. 3, pp. 15–64, 1993.
[6] G. Bierman, N. Benton, V. De Paiva, and M. Hyland, "A Term Calculus for Intuicionistic Linear Logic", *Computer Laboratory, University of Cambridge*, vol. 3, pp. 15–64, 1993.
[7] J. Y. Girard, "Linear logic: Its syntax and semantics", *Cambridge University Press*, vol. 3, pp. 15–64, 2003.
[8] J. Y. Girard, "Proofs and types", *Cambridge University Press*, vol. 3, pp. 15–64, 1990.
[9] V. Novitzká, D. Mihályi, and V. Slodičák, "Linear logical reasoning on programming", *Acta Electrotechnica et Informatica No.3*, vol. Vol.6, pp. 15–64, 2006.
[10] V. Novitzká and V. Slodičák, "Kategorické štruktúry a ich aplikácie v informatike", ISBN 978-80-89284-67-2, 2010.
[11] D. Turner, "Church's thesis and functional programming", *Middlesex University*, vol. 3, pp. 15–64, 2006.

Markerless augmented reality using electromagnetic tracking device

[1]*Martin VARGA (2st year),* [2]*Peter IVANČÁK (2st year)*
Supervisor: [3]*Branislav SOBOTA*

[1,2,3]Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

[1]martin.varga@tuke.sk, [2]peter.ivancak@tuke.sk, [3]branislav.sobota@tuke.sk

***Abstract*— This paper presents selected parts of markerless augmented reality application based on electromagnetic tracking device. Paper is divided into three parts. The first part is about augmented reality and its technology. The second part describes advantages and disadvantages of see-through and video-through head-mounted displays. This part also focuses on the determination of the position using electromagnetic tracking device. The last part of the paper describes the architecture of markerless augmented reality system using electromagnetic tracking device.**

***Keywords*—augmented reality, markerless augmented reality, head tracking, head-mounted display**

I. INTRODUCTION

There are several technologies applicable when creating VR (virtual reality) applications. One of these technologies is mixed reality. Mixed reality (MR) is an area of computer science research that aims at combination of real world with generated data (virtual reality). Such computer generated graphical objects are blended in to real environment in real time. Real world data are provided though sensors in real time back to mixed reality system [1], [7]. Mixed reality can according to [2] originate from at least one of these technologies: augmented reality (AR) and augmented virtuality (AV).

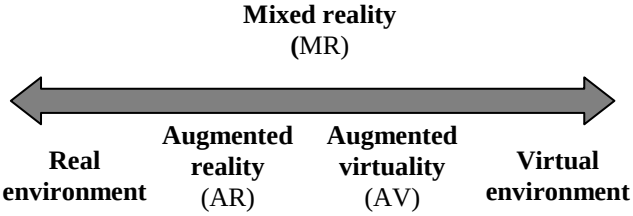


Fig. 1. Milgram's definition of real to virtual world transition (reality-virtually continuum)

Augmented Reality (AR) merges a real world and a virtual environment. A virtual object is added into a real world in order to improve or to add more information for an observer. AR is computer-generated data integration with the real world, which among others can be done with computer graphics rendering on a real-time footage. AR can be used for many things, such as displaying a mobile directions to head-up display, in the medical field, the AR may help doctors to insert information on a patient's medical record (such as x-ray result from the patients), or to reconstruct the old buildings and historic as reality which can be seen at present time. Virtual Reality (VR) was developed using a concept that actually use the environment which is engineered in such a way that resembles the real world [6]. More information about this technology can be found at [1]-[7] [11], [13].

According to the method how virtual objects are aligned with real scene image there are two systems use:

- *Marker systems* – special markers are used in a real scene. The markers are then recognized during runtime and replaced with virtual objects.
- *Markerless systems* – processing and inserting of virtual objects is without special markers. Additional information is needed, for example image (e.g. photo (semi-markerless system)), face recognition, GPS data, inertial and electromagnetic tracking devices, etc.

Augmented virtuality is technologically similar to augmented reality. However its principle is opposite to AR. Most of a world (scene) that is displayed is virtual and real objects are inserted in the world. If an observer is inserted into scene he or she is dynamically integrated in AV system in same manner as other inserted real objects. That allows real time manipulation with virtual and real objects inside scene. More information about this technology can be found at [9], [10].

Both augmented reality and augmented virtuality systems are quite similar and they belong to mixed reality definition. A goal of mixed reality system is to merge real world with virtual one into new environment where real and virtual (synthetic) objects exist together and interact in real time. Relationship between mixed reality, augmented reality and augmented virtuality is defined by Fig.1 [2], [3].

II. HEAD-MOUNTED DISPLAY AND ELECTROMAGNETIC TRACKING DEVICE

A head-mounted display (HMD) is display device, worn on the head or as part of a helmet that has a small display optic in front of one (monocular HMD) or each eye (binocular HMD).

Based on how a user sees mixed reality there can be two types of systems:

- *Optical see-through* systems where the user sees real world directly and computer generated objects are

added to this view. This category of systems usually works with semi-transparent displays.

- *Video see-through* where captured real world image with added virtual objects is displayed to the user. This is usually realized via camera – display system [4].

A. Optical see-through HMD

Advantages and disadvantages of the optical approach are:

1. Simplicity. Optical blending is much simpler and cheaper than video blending.
2. Resolution. Video blending limits the resolution of what the user sees. The user view of the real world is not degraded in the optical approach because the user is directly looking at the real world.
3. No eye offset. In the video see-through approach a camera is used to capture the real world. Since the cameras are not exactly at the same position as the human's eye, there is always a small offset. Compensating for this offset is possible. In the optical approach the user is looking directly at the real world and therefore has no eye offset [5].

B. Video see-through HMD

Advantages and disadvantages of the video approach are:

1. Flexibility. With the optical see-through it is difficult to completely obscure the real world objects with virtual objects. In the video see-through approach the real world is available as a digital stream. Therefore it is much easier to blend the two. Because of the flexibility, video see-through may ultimately procedure more compelling environments than optical see-through approaches.
2. Real and virtual view delays can be matched: Video offers an approach for reducing or avoiding problems caused by temporal mismatches between the real and virtual images. Optic al see-through HMDs offer an almost instantaneous view of the real world but a delayed of the virtual. This temporal mismatch can cause problems. With video approaches, it is possible to delay the video of real world to match the delay from the virtual image stream.
3. Additional registration strategies. In optical see-through, the only information the system has about the user's head location comes from the head tracker. Video blending provides another source of information: the digitized image of real scene. This digitized image means that video approaches can employ additional registration strategies unavailable to optical approaches [5].

C. Electromagnetic tracking device

Electromagnetic spatial measurement systems determine the location of objects that are embedded with sensor coils. When the object is placed inside controlled, varying magnetic fields, voltages are induced in the sensor coils. These induced voltages are used by the measurement system to calculate the position and orientation of the object.

Electromagnetic device is usually composed of two parts (Magnetic source and sensor).

Magnetic source -The Source is the device which produces the electro-magnetic field and is normally the reference for the position and orientation measurements of the sensors. It is usually mounted in a fixed position to a non-metallic surface or stand, which is located in close proximity to the sensors.

Sensor(s) - The sensor is the smaller device whose position and orientation is measured relative to the Source.

Those parts are used to calculate a position and orientation of the sensor in 3D space. The azimuth, elevation, and roll angles that define the current orientation of the sensor coordinate frame with respect to the designated reference frame.

The Euler angle coordinates that are output by for example Polhemus PATRIOT as one measure of sensor orientation are graphically defined in Fig. 2. Here, the x, y, z and X, Y, Z three-axis arrays represent independent, three-dimensional orthogonal coordinate frames. The x, y, z triad represents the sensor frame in its current orientation state. The X, Y, Z triad represents the reference frame against which the relative orientation of the sensor frame is measured. By definition, the X, Y, Z frame also represents the zero orientation reference state of the sensor frame.

The Euler angles, azimuth, elevation and roll, are designated ψ, θ, and φ. These angles represent an azimuth-primary sequence of frame rotations that define the current orientation of the sensor with respect to its zero orientation state. The defining rotation sequence is an azimuth rotation followed by an elevation rotation followed by a roll rotation.

The azimuth angle ψ is defined in the figure as a rotation of the X and Y reference axes about the Z reference axis. The transition axes labeled X' and Y' represent the orientation of the X and Y axes after the azimuth rotation.

The elevation angle θ is defined as a rotation of the Z reference axis and the X' transition axis about the Y' transition axis. The transition axis labeled Z' represents the orientation of the Z reference axis after the elevation rotation. The current x-axis of the current sensor frame represents the orientation of the X transition axis after the elevation rotation.

Lastly, the roll angle φ is defined as a rotation of the Y' and Z' transition axes about the x-axis of the sensor frame. The y and z-axes of the current sensor frame represent the orientation of the Y' and Z' transition axes after the roll rotation

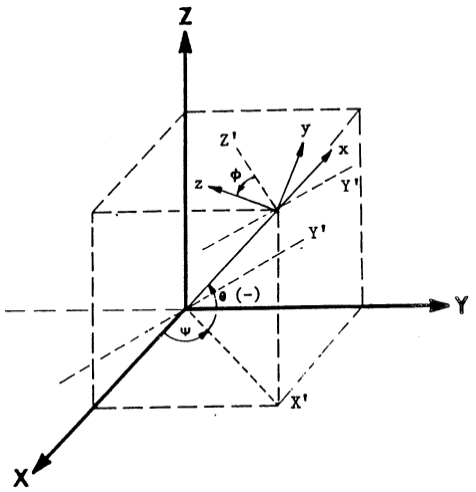


Fig. 2. Reference frame of system Polhemus Patriot (X,Y,Z - alignment (reference) frame, x,y,z – rotated sensor coordinate frame, ψ – Azimuth, θ – Elevation, φ - Roll)[12]

III. MARKERLESS AR SYSTEM USING ELECTROMAGNETIC TRACKING DEVICE

Markerless augmented reality consists of three main components: *Initialization*, *Head tracking* and *AR visualization*.

The component *Initialization* sets virtual 3D scene which was created using 3D models in OBJ format. The component *Head Tracking* acquires human head position and orientation using electromagnetic device. After that the component sets transform matrix of view (GL_MODELVIEW). The latest component (*AR visualization*) displays the virtual scene in the real world using head-mounted display.

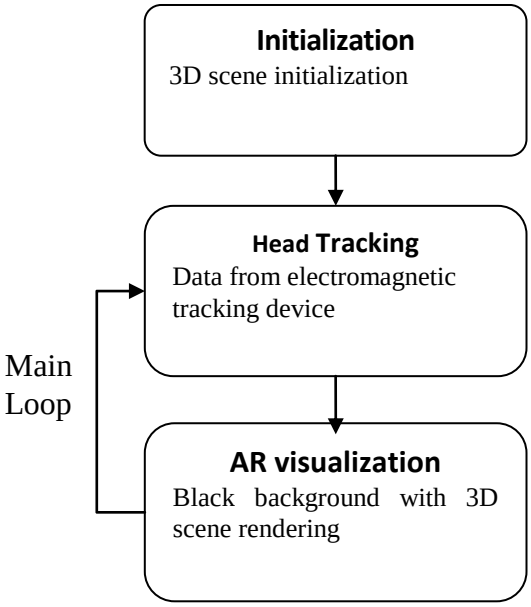


Fig. 3. The architecture of the markerless AR system using electromagnetic tracking device and the visualization of virtual scene (using head mounted display).

A. *Initialization*

This virtual scene is composed from 3D objects which were created in the stand-alone software. The user is able to locate 3D objects in the virtual scene and to create animation with those objects.

The last step is export of the XML document. This document contents the virtual scene and is loaded in the visualization application. Next step is a detection of human head for correct view of the virtual scene in a real environment.

B. *Head Tracking*

Polhemus PATRIOT (Fig. 4) is used for head tracking in our solution. PATRIOT provides dynamic, real-time measurements of head position (X, Y and Z Cartesian coordinates) and orientation (azimuth, elevation and roll). PATRIOT can update data continuously, discretely (point by point), or incrementally. Sensor is located on the top of head-mounted display and it measures correct position and orientation of user head. The origin of coordinates system of virtual scene is identical with magnetic source. Fig. 4 shows a schematic representation of a system for markerless augmented reality.

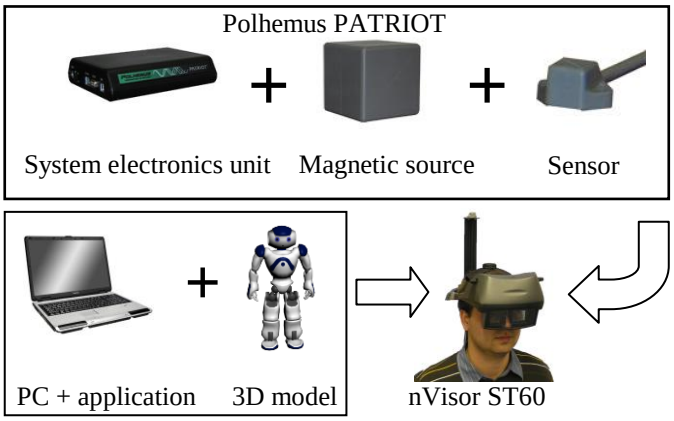


Fig. 4. The schematic representation of a system for augmented reality using a head tracking device and 3D visualization (using head-mounted display). Polhemus PATRIOT consists of three main parts (System electronics unit, magnetic source and sensor).

A. *AR visualization*

The system is implemented on MS Windows platform. Markerless augmented reality application uses 3D model in OBJ format and OpenGl libraries for rendering (Fig.3 section "AR visualization").

For displaying the head mounted-display nVisor ST60 was used (see Fig. 4, Fig. 5). This HMD use optical see-through technology to create illusion of three dimensional objects in the real world. For displaying is used Liquid crystal on silicon (LCOS) technology. Displaying resolution is 1280×1024. Weight of this HMD is 1300 g.

Rendered image of virtual scene created by application is displayed using HMD. Black background is not displayed and the user sees real world augmented with virtual scene (Fig. 5 on the right). To create stereoscopic vision, application renders two images which are side by side. One image is for the left eye and another one for the right eye.

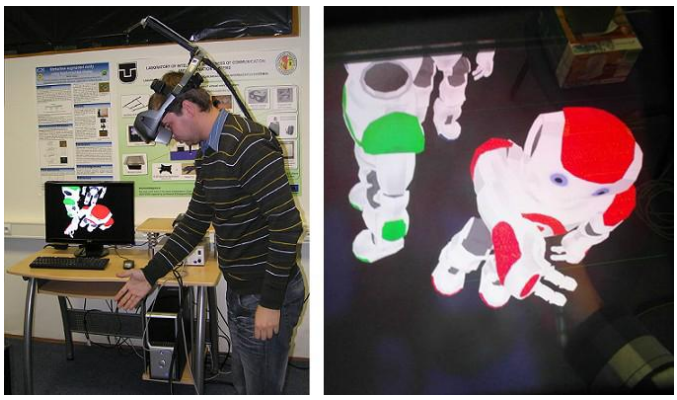


Fig. 5. The user works with AR system: The user with head mounted display is on the left and real user's view with added virtual 3D objects is on the right [8].

IV. CONCLUSION

The proposed system for markerless AR display, presented in the paper, was constructed at the Department of Computers and Informatics.

Advantage of this application is that the user doesn't need any marker for calculation of position. Application displays virtual object into user's view using HMD. User sees the real

world with virtual scene in the same view. Disadvantage is that the electromagnetic tracking device is interfered by metal items and magnetic field. Future work will be focused on the localization of user's head position and orientation with a 3D inertial motion tracker.

ACKNOWLEDGMENT

This work has been supported by KEGA grant project No. 050TUKE-4/2012: "Application of Virtual Reality Technologies in Teaching Formal Methods" and this work is also the result of the project implementation HUSK/1101/1.2.1/0039 "Virtual Reality Laboratory for Factory of the Future - VIRTLAB" supported by the Hungary – Slovakia Cross-Border Co-operation Programme 2007-2013 funded by the ERDF.

REFERENCES

[1] AZUMA, R., Tracking Requirements for Augmented Reality, Communications of the ACM Vol. 36, No. 7, 1993,pp. 50-51.

[2] MILGRAM, P., KISHINO, F., A Taxonomy of Mixed Reality Visual Displays, IEICE Transactions on Information Systems, Vol. E77-D, No. 12, 1994, pp. 1321-1329.

[3] HROZEK, F., SOBOTA, B. SZABÓ, Cs., KOREČKO, Š., VARGA, M., IVANČAK, P., Augmented reality application in parallel computing system, 7th International Workshop on Grid Computing for Complex Problems, Bratislava, Slovakia, 24 –26 October 2011,Ústav Informatiky SAV, 2011, pp.118-125,978-80-970145-5-1.

[4] SOBOTA, B., JANOŠO, R., 3D interface Based On Augmented Reality In Client Server Environment, Journal of information, control and management systems, Vol. 8, No. 3,2010, pp. 247 – 256, ISSN 1336-1716.

[5] AZUMA, R., A Survey of Augmented Reality, In presence: Teleoperators and Virtual Environments 6, 1997,pp. 355 – 385.

[6] ARIYANA, Y.; WURYABDARI, A.I., "Basic 3D interaction techniques in Augmented Reality," *System Engineering and Technology (ICSET), 2012 International Conference on* , vol., no., pp.1,6, 11-12 Sept. 2012

[7] SZABÓ, Cs., KOREČKO, Š., SOBOTA, B.: "Data Processing for Virtual Reality," In: Advances in Robotics and Virtual Reality, Intelligent Systems Reference Library Vol. 26, Springer Berlin-Heidelberg, 2012, pp. 333-361, ISBN 978-3-642-23362-3, ISSN 1868-4394.

[8] HOVANEC, M., VARGA, M., a kol., Inovatívne trendy a vízie v ergonómii využitím rozšírenej a virtuálnej reality. 2012.In: Aktuálne otázky bezpečnosti práce : 25. medzinárodná konferencia : Štrbské Pleso - Vysoké Tatry, 06.-08. 11.2012. - Košice : TU, 2012 S. 1-7. ISBN 978-80-553-1113-5.

[9] XIANGYU Wang, "Specifying Augmented Virtuality Systems for Creative Architectural Design," Information Visualization, 2007. IV '07. 11th International Conference , vol., no., pp.584,589, 4-6 July 2007

[10] CLARKE, J.; VINES, J.; MARK, E., "An augmented virtuality scientific data center," User Group Conference, 2003. Proceedings , vol., no., pp.354,357, 9-13 June 2003

[11] SOBOTA, B. – KOREČKO, Š. – LÁTKA, O. – Szabo, Cz. – HROZEK, F.: Riešenie úloh spracovania rozsiahlych grafických údajov v prostredí paralelných počítačových systémov. 2012, Editačné stredisko TU, Košice, ISBN 978-80-553-0864-7

[12] Polhemus Patriot, homepage [online], url: http://www.polhemus.com/?page=motion_patriot

[13] DOS SANTOS, A.L.; LEMOS, D.; LINDOSO, J.E.F.; TEICHRIEB, V.; , "Real Time Ray Tracing for Augmented Reality," Virtual and Augmented Reality (SVR), 2012 14th Symposium on , vol., no., pp.131-140, 28-31 May 2012

Modular Computer Language Process

[1]*Dominik LAKATOŠ (3[rd] year)*
Supervisor: [2]Jaroslav PORUBÄN

[1,2]Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

[1]dominik.lakatos@tuke.sk, [2]jaroslav.poruban@tuke.sk

***Abstract*—Specification of a new computer language and creation of required tools is a common problem. Tools are usually built for one specific language or in better case in form of generators for language workbenches. Existing generators of supporting tools for languages, such as lexers, parsers or editors are not exchangeable and requires own specification. In this paper we want to provide overview to this problem and possible solution through modularization of an entire language process.**

***Keywords*—computer language, language process, modularization, parser**

I. Introduction

Languages surround us in our everyday life and we need to understand them properly in order to communicate effectively with our surroundings. Each natural language corresponds to the environment in which it was created and evolved to current state, its grammar and meaning of words is influenced by the way the people are thinking. It is not easy to learn and understand language from distant countries as we don't know more about their culture and the way they live. Understanding the presence of significant differences in languages is the cornerstone for mastering any other non-native language. The human mind have to be able to process each language differently, almost like if there were separate processing parts for each language.

Actually we don't want to study processes in human brain concerning processing of languages, our idea is just to try and learn something from this process. People speaking different languages often need to "switch" from one language to other in order to speak and understand it better. It is as if there were some interchangeable parts in our mind specialized for each language. The same should apply to computer languages and processing of computer languages in machines. It is well-known that there already are different processors for each computer language, usually also called parsers. Those are used totally separately for processing of each language. In case that you need to process C language, you need specific C compiler, and for processing of Java language everything have to be replaced. Every language needs its own specific tools in order to be understood by computer. At least there is one common point for each language as almost every computer language (except the ones used to represent data structure) is at end translated to sequence of processor instructions. Still, it is discussable way of processing language as in comparison to the human brain; the computer processing is still much less adaptable and much more coarse-grained in way of having totally different tools for each language.

In the next parts we would like to present one improvement for processing languages, which can help a little in bringing computer language processing to different level. In order to archive it we propose usage of modular language process, every language processing part could be modular and replaceable to better fit needs of each computer language. Processing of languages should be more flexible.

II. Phases of language lifecycle

A. Language processing

Processing of computer language consists of several phases. The basic ones are lexical analysis, syntactic analysis and semantic analysis with later code generation and execution or interpretation [1].

Lexical analysis is a process for recognition language tokens from sequence of characters. There are many widely known tools for lexical analysis (lexers), such as Lex [2], Flex, JFlex [3], Quex [4], ANTLR [5], JavaCC [6]. Result of the lexical analysis is a sequence of tokens, which are later used in next phase.

Syntactical analysis, also called parsing, is the next phase in language processing. The tokens are checked if they form an allowable language expression or sentence and then they are converted to abstract syntax tree. Parsers could be implemented specifically for one language for efficiency purposes or it is possible to use parser (or compiler) generator tools, the common ones include Yacc [7], ANTLR [5], Bison [8], JavaCC [6], Beaver [9]. In most cases lexical analysis and syntactical analysis are communicating during language processing as languages could even have different techniques for processing different parts of sentences, as having different set of tokens acceptable in different states. MetaLexer [10] is a generator of lexical analysis tool with support for composing languages and defining different tokens for different parts of languages.

Semantic analysis is the final phase of language processing before preparing for execution. It can be defined as translation from one language to another (e.g. generating code) or interpretation of expressions. Attribute grammars are also usable to define these actions. Some of the actions usually included in semantic analysis consist of resolving textual identifiers and its references.

We think that reference resolving actions should be extracted from semantic analysis as they don't form actual

behavior of sentence; they are just transformations from abstract syntax tree to abstract syntax graph, which is structure with connected references to identifiers. This introduces new phase into language processing, that is reference resolving and it fits right after syntax analysis and before semantic analysis. Reference resolving phase is only needed for textual languages, as graphical languages usually don't need to use identifiers for connecting language elements together. Overview of discussed language processing phases and its inputs and outputs are displayed on Fig. 1.

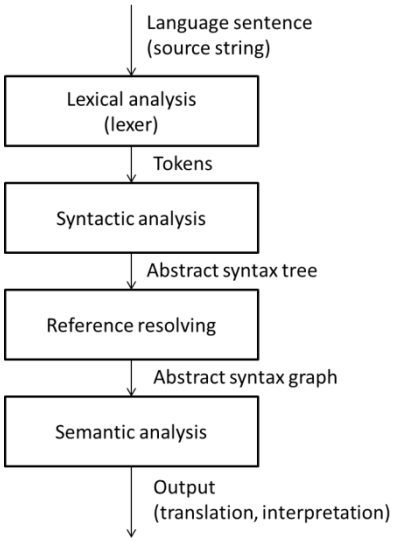


Fig. 1. Phases of language processing

Every existing tool for generating lexers, parsers and processing semantics is mostly based on its own specification language. This does not allow for tools to interoperate. It is not possible to choose any lexical analysis tool and any parsing tool and use the best possible combination of existing tools. The tools are not undependably pluggable into the language process.

B. *Specification of languages*

Processing of languages is not the only part of the language lifecycle. Before we can write sentences in language we need to create and specify its concepts, which can later be used for specification of syntax and semantics.

First computer languages were specified mostly in form of grammars. Abstract syntax of such languages were not defined explicitly, we could discover the concepts of each language only by analyzing the grammar rules and non-terminals. Even semantics were defined on non-terminal. This is still great solution for languages, which requires maximum performance at cost of adaptability.

Most new and modern computer languages, especially domain specific languages, require flexibility instead of performance. Flexible solution for defining computer languages is about simple way for writing specification, extensibility and reusability. We need such features mostly for domain-specific languages [11]. In such cases language designer would like to build language simply by reusing existing language components, such as language parts for variables, mathematical operations, etc.

Basic tools for parsing sentences of specified languages are not sufficient for modern languages. We need flexible languages with flexible and modern tools such as editor with syntax-highlighting, auto-completion, textual deserialization, etc [12]. Thanks to modern tools, as are various language workbenches(Xtext [13], MetaEdit+ [14], COPE [15]), it is possible to specify language and generate appropriate tools [16]. Options of each language workbench are closed only for languages specified within this tool, so every language workbench need to create its own set of generated language tools [17].

III. LANGUAGE MODELING

In previous sections we have described simple lifecycle of computer language, which should be first specified and then parser and other supporting tools would be generated. We want to build upon this lifecycle and modularize it.

First step in modularization is definition of one simple but still extensible language specification form. There are many specification languages and forms for designing language, some are textual and some are graphical. In our case we have chosen to model the language in form of classes used in object-oriented programming. Such classes can easily represent concepts of languages and have both textual and graphical representation. Textual representation is classical implementation of class in any programming language and graphical representation is in form of UML class diagram. Every language concept can be easily extended with usage of generalization feature of object-oriented programming. For storing additional information required for each concept we can use annotations on any class member. Set of annotations is not restricted and can also be expanded. Details about specifying new language in this form are explained in [18] [19].

When we have one form for specification of languages, we can use information contained in it to generate almost all tools, as long as tool specific required information is provided. It is possible to analyze designed language and recommend the best parser and language tools to be generated. Language designer should not need to care about proper parsing method (LL, LR, LALR, etc.) for his language or any other problems concerning language processing as long as the language is properly designed. Languages designed in different specification form can be transformed to the object-oriented format in order to unify language design specification.

IV. MODULAR LANGUAGE PROCESS

Modularization is not completely new trend as Hudak [20] proposes to modularize description of language, but it is different from our perspective of modularization of tools generated for language process. In order to support our claims about modularization of language lifecycle and processing of language in previous sections of this paper, we have designed and tested system for language design. Basis of our modular design is YAJCo tool [18] with its object-oriented classes for specification of language concepts.

A. *Modular generation of language parser and tools*

Language is defined through language concepts using notations of YAJCo tool for defining language structure and concrete syntax. There is also support for defining identifiers and textual references to existing identifiers. It supports even

other common language design patterns as operator pattern for notating language concepts as operators with one or 2 operands and it allows to specify priority and associativity of operation.

Generation of language tools in our language process consists of three main steps, which are (a) reading the language model, (b) generating parser and (c) generating language tools as it is depicted on Fig. 2.

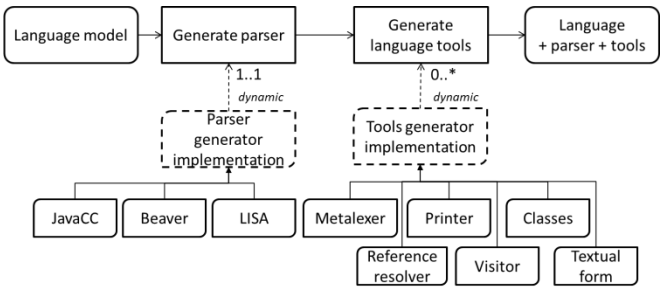


Fig. 2. Modular process of generating language parser and tools

Language model can be specified in form of (a) Java classes, (b) textual form, or (c) created dynamically in a program as instances of language model concepts. After we have language model, we can validate and analyze it for recommendation of the best parsing technology and suitability for specific tools. Every tool generator can provide its own validation actions. The main advantage of our solution is simple process without any special dependencies on actual implementations for generators of parser or language tools. The process is simply directing the generation, but actual generators are discovered dynamically during runtime. Every generator is required to implement only one simple interface with method for generating files:

```
generateFiles(Language, Filer, Properties)
```

It takes 3 parameters: *Language* is model of designed language, *Filer* is class responsible for providing access to files and marking new source files for compilation in the end of generating process and *Properties* is used as storage of options, which can be used to customize generators. Every generator can internally specify own options and then use global configuration support of our process to read all options from one place or even combine them from multiple places (annotations, Java property file or XML file).

If you want to use any of implemented generator tools the only needed action is including the tool into libraries and it is automatically discovered and used. Some tools can have dependence on other existing tools, e.g. generated source use source generated from other tool. In this case it is possible to specify the dependency with simple annotation:

```
@DependsOn("yajco.generator.Visitor")
```

Parser generator implementation used during modular process is only specially marked implementation of the tool generator. Special notation is used for checking for exactly one parser generator, as we support declaration of parser class name in designed language and generation of more than one parser would not be supported in this scenario, but it is only configuration restriction and can be changed.

Implementation of different parser generators for our process allows not only effective changing of parsing technique but can have other advantages as possibility to use other tools included in parser generator, e.g. LISA parser generator [21] has step-based graphical animation of parsing language sentence, which can be used by students to better understand parsing techniques.

B. Modular language processing

Modularization of generator tools is only one of possible places to include modular functionality. The same method can be used in actual parsing process. Every parsing phase can consist of one module. We only need to define proper interfaces for each phase and change implementation of generators to reflect changes. Even phases as lexical analysis and syntactical analysis, which are mostly served directly by generated parsers, can be separated. In order to support composition of languages we have implemented generator for lexical specification language Metalexer [10], which support composition and extension of lexical specification. Such lexer can be dynamically discovered and used instead of lexical analyzer included in parser.

The same applies to phase of resolving references and creating abstract syntax graph from abstract syntax graph. For resolving references there are actually two possible options. One requires changing of semantics actions used in parser generators, that means it is not very modular for language processing phase. The second type of our reference resolver is based on aspect programming methods and involves registration of each instantiation of language concept to special tree structure, which is used later for actual resolutions of textual references and producing of abstract syntax graph. Final representation of abstract syntax tree is in form of instances of language concepts.

Semantic analysis for any language designed with our method is simple enough as it can use actual language concept objects. Each object represents one concept and actual semantics actions can be included in separate module or included directly in classes of language concepts.

V. Conclusion

This paper serves as simple overview of modular computer language process. We have introduced method for making this process independent from any implementation. Every phase of the process can be modularized and implementations made in form of pluggable libraries. Every language tool generator or parser generator can be used in our process as long as transformation is defined or discovered between specification in YAJCo tool and required third-party tool.

For purpose of evaluating this method, we have defined interfaces for basic modules and implemented transformations between YAJCo tool and three different parser generators, one special featured lexer, reference resolver based on AspectJ technology as well as few other language tools generators. The method has been tested on couple languages with success and seems promising for future usage. Still there is a lot of work ahead to make the modularity of process complete, mostly we would need to better modularize language processing phase with better interoperability between different kinds of lexers, parsers, reference resolvers and support different kinds of semantic analysis.

ACKNOWLEDGMENT

This work was supported by VEGA Grant No. 1/0305/11 – Co-evolution of the Artifacts Written in Domain-specific Languages Driven by Language Evolution.

REFERENCES

[1] K. C. Louden, Compiler Construction: Principles and Practice, Boston, MA, USA: PWS Publishing Co., 1997.

[2] M. E. Lesk and E. Schmidt, Lex: A lexical analyzer generator, Murray Hill, NJ: Bell Laboratories, 1975.

[3] G. Klein, "JFlex: The fast scanner generator for Java," [Online]. Available: http://jflex.de. [Accessed 9 3 2013].

[4] F.-R. Schäfer, "Quex - Fast Universal Lexical Analyzer Generator," [Online]. Available: http://quex.sourceforge.net/. [Accessed 10 March 2013].

[5] T. J. Parr and R. W. Quong, "ANTLR: A predicated-LL(k) parser generator," *Software: Practice and Experience,* vol. 25, no. 7, pp. 789-810, 1995.

[6] V. Kodaganallur, "Incorporating Language Processing into Java Applications: A JavaCC Tutorial," *IEEE Softw.*, vol. 21, no. 4, pp. 70-77, July 2004.

[7] S. C. Johnson, Yacc: Yet another compiler-compiler, Murray Hill, NJ: Bell Laboratories, 1978.

[8] C. Donnelly and R. Stallman, "Bison: The YACC-compatible Parser Generator," 1995. [Online]. Available: http://dinosaur.compilertools.net/bison/index.html.

[9] A. Demenchuk, "Beaver - a LALR Parser Generator," [Online]. Available: http://beaver.sourceforge.net/. [Accessed 8 March 2013].

[10] A. Casey and L. Hendren, "MetaLexer: a modular lexical specification language," in *Proceedings of the tenth international conference on Aspect-oriented software development*, Porto de Galinhas, Brazil, 2011.

[11] M. Fowler, Domain-Specific Languages, Addison-Wesley Professional, 2010.

[12] D. Lakatoš and J. Porubän, "Generating Tools from a Computer Language Definition," in *Proceedings of International Scientific conference on Computer Science and Engineering*, Košice - Stará Ľubovňa, Slovakia, 2010.

[13] "Xtext," [Online]. Available: http://www.eclipse.org/Xtext/. [Accessed 11 March 2013].

[14] "MetaEdit+," [Online]. Available: http://www.metacase.com. [Accessed 3 March 2013].

[15] M. Herrmannsdoerfer, "COPE - A Workbench for the Coupled Evolution of Metamodels and Models," *Software Language Engineering*, vol. 6563, pp. 286-295, 2011.

[16] M. Fowler, "Language Workbenches: The Killer-App for Domain Specific Languages?," June 2005. [Online]. Available: http://www.martinfowler.com/articles/languageWorkbench.html. [Accessed 6 March 2013].

[17] M. Pfeiffer and J. Pichler, "A Comparison of Tool Support for Textual Domain-Specific Languages," *The 8th OOPSLA Workshop on Domain-Specific Modeling*, pp. 1-7, 2008.

[18] J. Porubän, M. Forgáč, M. Sabo and M. Běhálek, "Annotation Based Parser Generator," *Computer Science and Information Systems,* vol. 7, no. 2, pp. 291-307, 2010.

[19] D. Lakatoš, J. Porubän and M. Sabo, "Assisted software language creation using internal model," *Journal of Computer Science and Control Systems,* vol. 4, no. 1, pp. 71-74, 2011.

[20] P. Hudak, "Modular Domain Specific Languages and Tools," in *Proceedings of the 5th International Conference on Software Reuse*, Washington, DC, USA, 1998.

[21] M. Mernik, N. Korbar and V. Žumer, "LISA: a tool for automatic language implementation," *SIGPLAN Not.*, vol. 30, no. 4, pp. 71-79, April 1995.

Monitoring of critical issues in departmental IT architecture

[1]*Roman MIHAĽ (3rd year),* [2]*Peter KARCH*
Supervisor: [3]*Iveta ZOLOTOVÁ*

[1,2,3] Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

[1]roman.mihal@tuke.sk, [2]peter.karch@tuke.sk, [3]iveta.zolotova@tuke.sk

***Abstract*—This paper presents concept of monitoring system to help analyze and monitor the states of devices, services, computers or virtual computers included in IT network of organization. This article focuses on application of monitoring system in the departmental IT infrastructure environment with the computers and also virtual computers which act as a single computer nodes in the network architecture.**

***Keywords*—monitoring, network, OTRS, virtual computers, architecture**

I. Introduction

There are usually some responsive people and staff in the background of functional computers, networks, other connected devices like printers, and services provided by business in organization. But if we consider the computer network as a service producing environment for customers, being the administrators of systems or networks is then the key role in a business.

The maintenance of servers, computers, networks and software issues is very important to give a value to the customers for the money they have spent for.

Monitoring systems gives more effectiveness to the management of computer networks. These systems provide analyzing and controlling of the network or computers or running services (software instances) in periodic intervals.

The supplement to the function of monitoring actions is a system that classifies events in categories for the responsive administrator to give information about observed processes.

II. Systems Intended For Monitoring

A. Monitoring systems

Workstations, servers, printers and other devices are interconnected in the organization's infrastructure to simplify and accelerate the work results of employees, sizes them, and adjusts the resolution settings. Employees have created the access to servers for data retrieving or they have configured shared printers to print documents in effective way or they have other benefits provided by using computer networks.

If there is connected a large amount of networking devices and computers usually the rule is: the more networking heterogeneous devices you have included in your network, the greater risk of failures in established communication. During the outages in network it is necessary to solve the whole spectrum of reasons why the error occurred and where is the origin of the error.

If administrators of network system have prepared the documentation and it is updated in a regular base through the experiences with solved problems or events, the problem to find the error origin is minimized [1]. But there is other restraint in a term called time. The administrators usually get the information about problem from employees and then they try hard to fix these errors.

Sometimes or we can also say often there is a big gap between the time of commencement of error and the time of finding out about the error by employees who give the feedback to the administrators to restore the service, generally the previous state.

On the grounds of mentioned conditions experienced administrators decided to develop system which periodically monitor computer network and affiliated objects. At the moment when malfunction happen (e.g. it can be server outage, interruption of internet connection, unavailability of database communication, software services and so on) the monitoring system sends a notification of event to the administrator of the network. Afterwards the administrator can solve the problem.

B. Task oriented systems

The business sphere requires a powerful solution for effective communication between services provider, resellers of that services or products, and customers who are the most important in this circle. The customers buy goods and often there is a situation that due to incorrect using or manipulating with these goods they give questions to the responsive people from sales.

Analogically, there is the same situation with employees and administrators as responsive people in organization [2]. That big number of questions in a form of event are given to the administrators. Offered solution for this environment is a task system or coupon system or called like ticketing service which supports in a big way quick and effective execution of problem.

For example, the administrator of computer network in

organization arrived to the workplace and open email client and read messages. The manager needs to download a presentation file from a server but he can't sign in to the server with his credentials. The duty of administrator is to fix that problem however he just don't know [3]:

- where is the cause of error,
- whether it is non-functional only FTP service on server, or the whole server system,
- if this is problem of hardware or software character,
- whether the connection problem is only one-person malfunction or the other users (employees) share the same experiences,
- if the backup server is ready to represent functionality of original server.

III. Monitoring in Action

The monitoring system as mentioned above is a system that helps to analyze and find the source of problem so it gives a space to get know about problem in a more effective way. This approach is applicable in different organizations of various size. We decided to try specific features of monitoring system at our workplace, The Technical University of Košice, specifically in our laboratories at The Department of Cybernetics and Artificial Intelligence.

The laboratories are focused on technologies in the field of Cybernetics and Business Intelligence. The applications of the technologies run on a few servers which have virtual architecture consisting of hosts (act as a physical computer) and guests (virtual machines, virtual computers emulated on physical host computers). There are also computers with important role for data acquisition and to provide running state of technological models.

All these computers or virtual machines execute functionality permanently. The maintenance of laboratories is quite difficult so we decided for this test environment to determine results of implementing monitoring system.

The reasons to implement monitoring system:

- the most important reason to application of monitoring system is prevention of outages of observed objects. In the moment when interruption or collision in the computer network happen, the administrator is immediately acquainted with that problem. So the users of system in collision need not report behavior and errors to administrator and need not describe the situation. The administrator get a message about what is wrong with which monitored object it is not ok and with additional information. The time needed to identify what and where happened is slowing down using this solution which gives also more effectiveness to business and more continuity to the working process.
- using monitoring system emerges experiences with the same problems and in the next it is reasonable to change the network infrastructure or change conditions of running software instances. Nagios enables to monitor reaction of observed objects so it helps to stabilize the network or computational environment producing services. Nagios is also suitable tool for network testing with the aim to optimize it.
- Nagios provides also functionality to overview the structure of monitored objects through the internet browser. So it helps rather to understanding of network infrastructure and to preview the relations between nodes, devices, computers and so on. It is also possible to monitor system load and to predict some special cases preventing from complications.

IV. Application Concept

We decided to choose our laboratories at the Dept. of Cybernetics and Artificial Intelligence as an universal testing environment to understand in what way the use of monitoring system can help control applications, software instances, computers, virtual computers, etc. There are also technological models present in the lab with computer nodes representing data acquisition stations from these models.

As a software to ensure monitoring of devices and processes in our laboratories we choose OTRS (Open-source Ticket Request System) and a tool Nagios (Fig. 1).

The main priority of the software Nagios is monitoring of services and devices and in a real-time conditions informing authorized people about the network status or help in looking for solutions [4]. Nagios provides the scheme of devices connections in a graphical way therefore it helps during making documentation and supports understanding of network topology [5]. This tool is powerful and gives a possibility to monitor various types of problems.

In our conditions we created monitoring of:

- computers (or virtual computers) with installed OS

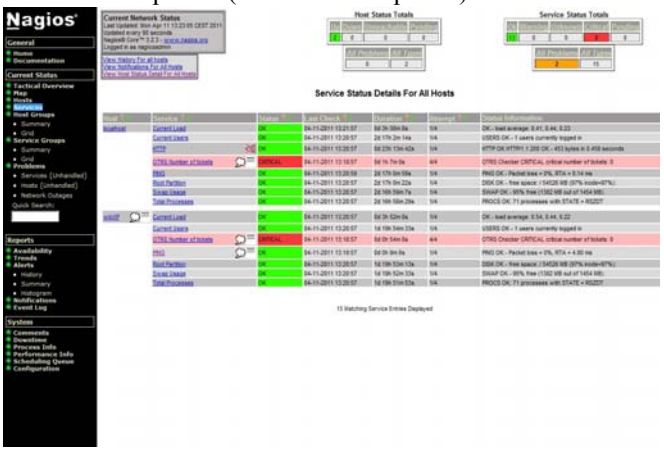


Fig. 1. Monitoring system Nagios

Windows – the monitoring function is based on external modules which send results into the main module of Nagios tool. There is installed NSClient++ agent to control the system information like status of RAM memory, amount of storage on disks, processor load, etc (Fig. 2).

- computers (or virtual computers) with installed OS Linux/Unix – monitoring configuration is a bit different from that with installed Windows, here it is based mostly on NRPE component.
- network printer – it is necessary that printer supports JetDirect protocol and has active protocol

SNMP.

- services – periodic check of services like SMTP, POP3, HTTP, SSH, etc.

The returned status from monitoring in Nagios can be in 3

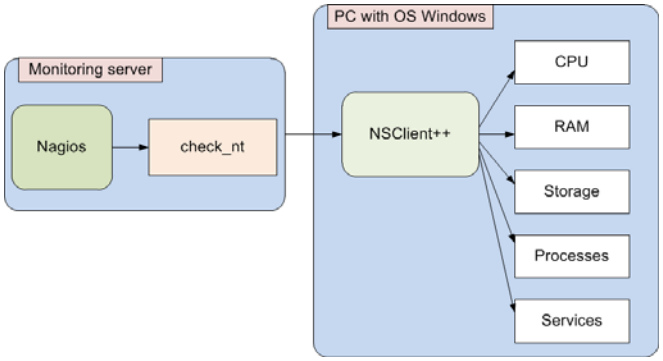


Fig. 2. Monitoring of computer with installed OS Windows

modes: OK, Warning, or Critical. The component which discover the error, sends message about status to Nagios and then it shows only one mode in the web interface from mentioned three modes.

In the combination with Nagios we decided to use OTRS (Open-source ticket Request System). OTRS (Fig. 3) provides a solution for effective information management in organization. It has three roles or views – the group of agents, group of customers and administrators group. Through the OTRS we can get a function of automatic generation of email in specific situation.

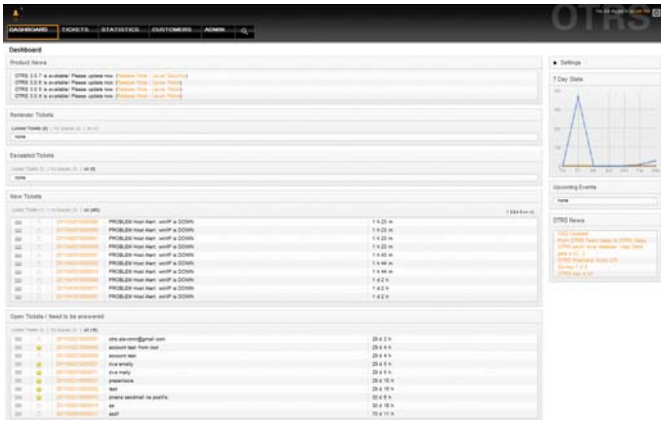


Fig. 3. OTRS web interface

The cooperation of both systems, Nagios and OTRS, is very important. In pursuance of created objects in Nagios, which represent particular monitored services, devices, computers, processes the Nagios sends an email at the detection of malfunctioning service [6]. On the other hand the OTRS can create tickets from incoming emails so it is possible to create a message for agent and this information is in the agents profile depicted like a ticket that is necessary to be solved by the administrator [7].

V. Conclusion

We describe a concept of monitoring system with a ticket service for administrators to let them know about actual problems and errors that have to be fixed. The testing environment are laboratories with computers, technological devices and virtual computers. The results are in more effective response to the problems in the labs so running systems are more reliable and have more stable behavior.

Our aim is to expand this system to the whole departmental level to be all technological systems, models or computers monitored and the administrators use a quick response on newly-emerged problems. We are preparing also to apply even best practices from Incident management as a part of ITIL framework.

Acknowledgment

This work was supported by grant KEGA No. 021TUKE-4/2012 (70%) and VEGA - 1/0286/11 (30%) project.

References

[1] S. Jarinčík, *Analýza a realizácia OTRS systému vo virtuálnom prostredí v spolupráci s monitorovacím systémom Nagios*. FEI TUKE, 2011, pp. 22.

[2] W. Barth, *Nagios, System and Network monitoring*. Open Source Press GmbH, 2005, ISBN 3-937514-09-0

[3] T. Scherbaum, *Praxisbuch NAGIOS*. O'Reilly Verlag GmbH & CO. KG., 2009, ISBN 978-3-89721-880-2

[4] J. Bothe, *OTRS - Nagios Integration*. NETWAYS Nagios Konferenz Nürnberg, 2008

[5] R. Mihaľ, M. Pavlík, P. Karch, *Alarms in Supervisory Control*. SCYR 2011 - 11th Scientific Conference of Young Researchers – FEI TU of Košice, 2011

[6] R. Mihaľ, *Incident manažment ako podpora rozhodovania a riadenia v systémoch SCADA/HMI*. Písomná práca k dizertačnej skúške, FEI TUKE, 2012

[7] M. Tvrdíková, *Aplikace moderních informačních technologií v řízení firmy*, *Nástroje ke zvyšování kvality informačních systémů*. Praha Grada Publishing, 2008. ISBN 978-80-247-2728-8

Monitoring of Employees by using monitoring tools

Ján ŠTOFA (2nd year), Peter MICHALIK (1st year)
Supervisor: Iveta ZOLOTOVÁ

Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

jan.stofa@tuke.sk, peter.michalik.2@tuke.sk, iveta.zolotova@tuke.sk

***Abstract*—This paper provides entrance into the issue of monitoring employees. Justifies its importance and compares the rights of employees and employers which are often in conflict. It draws attention the employee as a risk factor, who can endanger the safety of the company. The paper is an introduction for solving the problems of selecting suitable monitoring software with regard to the valid legislation of protection of personal data.**

***Keywords*—privacy, monitoring of employee, monitoring tools.**

I. INTRODUCTION

With the development of technology a growing range of tools, which allow monitoring the physical presence of the worker and his activity currently dealt with, also arises. The most recent example is on-line monitoring. While previously simple systems of "*time marker*" provided just information about presence of employees at the workplace, today more sophisticated systems can technically/ automatically monitor all content of their work. During this activity is very important do it carefully and do not violate the human rights on privacy or any other rights such as the secrecy of letters, communication and so on [4], [20].

The question of where is the line between electronic employee privacy and the protection of the interests of the employer is very sensitive topic. The number of cases of internal IT crime increases by misusage of assets of employers for private purposes. This has an impact on functionality and prestige of companies. Is monitoring the "*persecution*" of employee or multifunctional instrument providing feedback of personal, safety and preventive and budgetary policy of business entities [10]? The answer to this question is necessary to find through analysis of collected data. To be able to draw relevant results of our analysis we have to correctly answer next questions: Why do we need to monitor employees? An employee may be a risk factor for the company? What role on these issues plays legislation? Where to extend possibilities of monitoring software?

II. WHY TO MONITOR EMPLOYEES

One of the results of a survey conducted by Cisco dealt with the frustration of employees using a variety of devices and applications. This includes restrictions which IT managers specified on technology of cooperation. This finding reflects the fact that more than half of IT managers stated that in their company is prohibited use social networking applications. In some cases, employees take control into their own hands – half of respondents reported the use of prohibited applications once a week. More than half (52%) of organizations not allow usage of social networking sites or other similar applications at work [18].

Employee (external or internal) uses information and communication technology (ICT) as an authorized user. By using ICT employee ensures his working activities, which result from his working position. Hereby becomes a risk factor for the safe and efficient operation of ICT [21].

Surveys and practical experience of management, system administrators and consultants point to the fact that in area of working discipline in the use of computer equipment are habits of employees considerably deformed. From the perspective of employers are often undesirable habits. These facts are illustrated in the mentioned survey conducted in 2011 which include a total of 46 municipal offices of the Slovak Republic (Fig.1 and Fig.2) [10], [24].

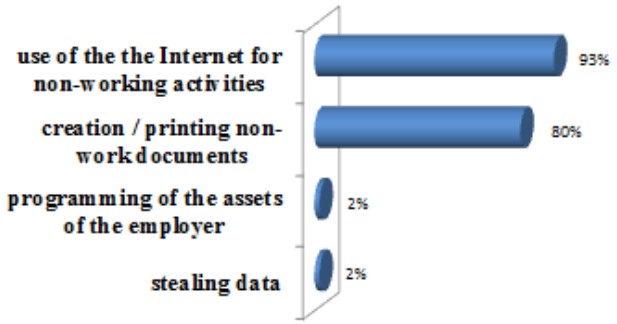


Fig. 1 The most commonly identified form of ICT use for non-working activities [10], [24]

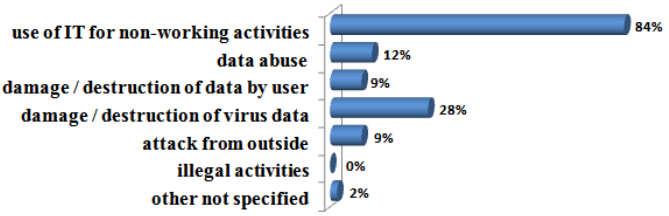


Fig. 2 The most commonly identified risk form of ICT use [10], [24]

III. EMPLOYEE AS RISK FACTOR

Undisciplined employee represents a major safety risk factor. He does not accept the fundamental requirements for his work with ICT. The current problematic phenomenon in the internal environment becomes increasing number

of cases of risk behaviour of users who interfere into criminal law. Management is often facing with the problem of the use of their assets. These employees are used to commit a crime:

- usage of ICT for private non-working activities:
  - browsing the Web unrelated to the work tasks
  - printing non-working data
  - burn/duplication of media
- download/dissemination of data subjected to copyright protection - SW / movie / music
- download/dissemination of data of immoral or illegal nature - xenophobia, racism, child pornography
- creation of non-authentic data to give these as authentic - documents to commit fraud of economic nature (tax, insurance, subsidy and various other)
- damage of the employer interests in various forms:
  - engaging in various not working discussions associated with ambiguous identity discussing separation from the employer
  - abuse of internet connectivity to mask the identity of the employer when posting on chat portals - particularly contributions with character of defamation, publication private compromising photos and videos of misleading adverts
  - disclosure of confidential data on discussion forums (such as information forming a trade secret of the employer, operational information)
- stealing data of employer to gain an advantage for:
  - its business
  - the purposes competitors
  - the other deliberate actions leading to disruption of the confidentiality, availability and integrity of data in order to harm the employer.

Inadequate employee behaviour against the assets of his employer (company) is drawn on following scheme (Fig.3) [3], [4], [10].

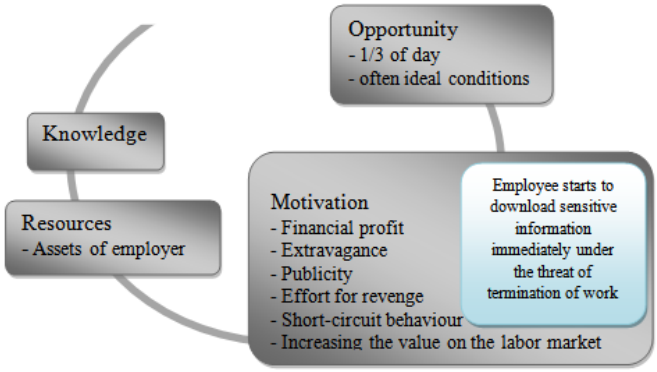


Fig. 3 Inadequate employee behaviour against the assets of employer [10]

IV. Legislation and Problem of Monitoring

Opinions on the theme legislation of monitoring are different. Into confrontation are:

- Employee's right of privacy
- Employer's right to control his employee

All of this legislation can be found in:

- § 10 paragraph 7 of Law no. 428/2002 Coll. Personal Data Protection
- § 12 of the Act. 40/1964 Coll. Civil Code
- Labour Code

Legislative treatment of this state in topic on-line monitoring is currently far from perfect. This will require the introduction of a clearly defined legal standard which will treat controlling of employee's activities in employment relationship [11], [23].

Labour Code solves this dilemma as King Solomon. The primary article provides that employer must not, without serious reasons to interfere employee's privacy in the workplace. The employer must not monitor employees without being notified and must not check the correspondence address to the employee as a private person. If a control mechanism is established, the employer must inform the employee about the range of control and its methods of implementation. It is not possible to monitor employees without the existence of a serious reason. Although this fact there is no statutory definition of what is a serious reason. General requirement is that the compelling reason has to be interpreted in direct response to the specific activity of the employer [11], [23].

Frequent problem is the introduction of "*secret*" monitoring forms without notifying the employees about the introduction of this form of control. This form of monitoring in the legislation of the Slovak Republic is unacceptable. Practical experience proves that the use of directives which inform about introduction of this form of control of the use of ICT, changes the behaviour of some of the users [22]. Novelty in addressing issue of monitoring would be to compare the existing monitoring software and their specific applications with existing legislative provisions.

V. Monitoring Tools

From the perspective of IT infrastructure, monitoring tools are exciting, and do not pursue primary work ethic, but also provide information on the use of systems and software applications in the enterprise. Their aim is to detect the intensity of the work of users with the monitored systems and applications. Consequently, it is possible to optimize the portfolio of hardware and the number of software licenses.

On the market there are several monitoring software systems. All software is specific by its functions that can be evaluated and compared to each other. We underline these monitoring software which meet important analyzed criteria: SONAR [7], Spector CNE Investigator [15], Spytech NetVizor [8], Imonitor EAM [12], Net Spy Pro [6], Work Examiner Standard [9], OsMonitor [13], REFOG Monitor [14], Activity Monitor [1], StaffCop [16], Yaware [17].

Analysis is divided into six monitored areas which particular software can do and what functions it provides:

1. **Online Monitoring**
   - Visited Websites: Records a history of the user's visited websites and lists the duration of each visit.
   - On-line Searches: User enters words into a search engine to locate websites that contain related content.
   - Chat: The software records the transcripts of both sides of chat and instant messaging conversations.
   - Gaming, Social web, YouTube, Sent/Received Email
   - Blogging: The purpose of a blog is to serve as vehicle for making online diary entries, commentary, thoughts about world happenings and more.
   - Usernames & Passwords: The combination of a U&P

is required to access some email, social networking websites and other accounts.
- File Transfers and Email Attachments: The software tracks all FT by the users of the monitored computer via email, external media, FTP and other methods.

2. PC Tracking & Management
- Real-Time Monitoring: PC administrators can view what is occurring on the monitored PC as it occurs.
- Display Screenshots: The spouse monitoring software records images of the monitored user's desktop.
- User PC Activities: Refers the activity that user performs on a monitored PC. Include accessed websites, launched applications and email exchanges.
- Key logger: Records every key stroke the user makes.
- Print Tracking: Tracks all printed documents.
- Document Tracking: Logs all files created, deleted and modified by users of the monitored computer.
- Portable Drives: The software tracks all external connected via USB ports or other methods.

3. Filtering & Blocking
- Website Filtering & Category Blocking: Software blocks internet content by subject categories.
- URL Black & White Lists: Allows or denies access to websites bases on administrator specified addresses.
- Keyword Blocking: The application blocks content that contains designated keywords.
- IM/Chat Blocking: Can block IM/Chat conversations based on application type, username or nickname.
- Application & Port Drive Blocking, Online Searches.

4. Reporting Methods
- Remote: The software delivers reports of the captured data to a designated email or FTP address.
- Built In Reports: The software has prebuilt reports.
- Local Network: Computer administrators can access the collected data across network shares.
- Management Console: PC admin can access collected data with a console interface installed on their PC.
- Top Ten Reports: provide access to the most frequently accessed websites, printers, searches…
- Screenshot Playback: The software plays back all captured desktop displays as if it were a video.
- Email Alerts: The application generates a report of all monitored user activity to designated email account.
- Custom Reports: PC administrators can create and store reports based upon their personal preferences.
- Search Capabilities
- Database: A database, hosted by an employer's server, houses the collected data and images.
- Data Retention: Computer administrators can specify how long data is saved and where on the network or local machine the information is stored.
- Graphical Charts
- Emailed: The reports of all computer use and online activities are sent to a designated email address.

5. Licensing & Support
- Included licenses: This term refers to the number of computers that can legally run the purchased SW.
- Upgrades: refer to a SW update. Some upgrades are included with the original product purchase for no additional cost.
- Tech Support

6. Operating Systems [1], [5], [6], [7], [8], [9], [12], [13], [14], [15], [16], [17].

Following chart represents results of scoreboard of analyzed monitoring software (Fig.4).

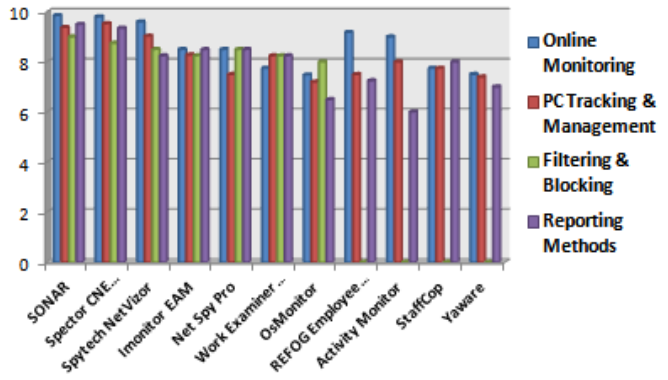


Fig. 4 Scoreboard of analyzed monitoring software [5]

For illustration we show the principle of operation of monitoring software (Fig.5).

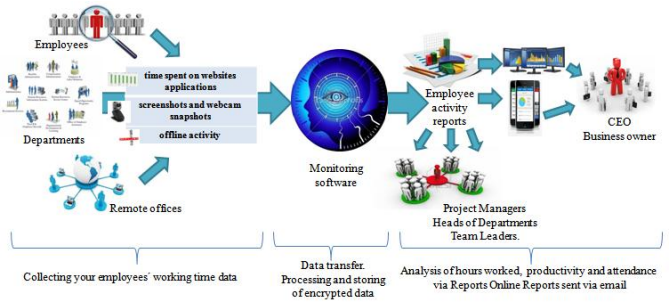


Fig. 5 Example of principle of operation of monitoring tools

Monitoring tools records data about the websites, programs, applications used by a user at the computer. Data is processed and presented as reports: by time (hours, days, weeks); for a whole team, by department or any individual employee. The obtained information helps the CEOs, managers, heads of departments and all other employees making better decisions regarding the use of working time and increases the overall efficiency of company work (Fig.6) [17].

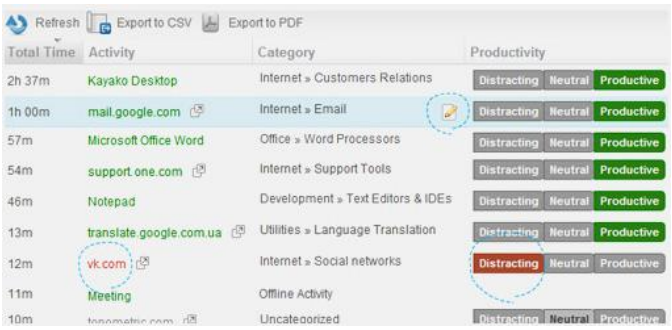


Fig. 6 Automated tracking of websites, example of monitoring tool Yaware [17]

Tasks of monitoring software is detailed monitoring of all or specific activities of computer users. In this way it is possible to monitor and evaluate the behaviour of employees

during their working hours. Characteristic features of these software features include:

- creating "*PrintScreen*" of computer screen
- recording any on-line communication including instant messaging protocols
- scanning characters entered from the keyboard
- history of surfing and searching on the internet
- activity monitor programs running on the monitored PC [2].

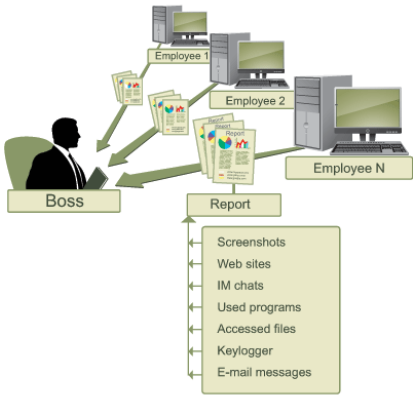


Fig. 7 Possibilities to process of monitoring data [21]

Scanned data and information is subsequently possible to review or analyze whenever is necessary (Fig.7). Based on data and info collected, the company is able to develop a detailed analysis. Through these we can create different models of processes comprising rules of behaviour of employees. Correctly modeled business processes will help safely and effectively use time of their employees [19], [21], [22].

VI. Conclusion

Trend in monitoring of employees is clear. Employers are trying to maximize the efficiency of their employees. Current leader in the forms of exploitation of working hours are mainly on instant messaging tools such as Skype and social networks in particular, represented by a network of Facebook. Many employers are afraid that their employees share on Facebook a lot of personal information. There are methods called social engineering that can elicit sensitive information and passwords. Another risk is caused by employee's negligence in manipulations with specific assets of the company. Reckless or intentional action of employees can cause a threat to the security breach of the company. In this regard we can see the efforts of employers to monitor work activities of their employees. Use of outcomes of monitoring tools can help companies in finding cases of failure among employees [4], [22].

Currently we are trying to obtain data from small and medium-sized enterprises. Thanks to them will be possible to respond in the introduction mentioned issues in greater detail. Based on the obtained data, we aim to clarify and choose the best company approached specific monitoring software. An important step is the introduction of measures such as monitoring workstations. Experience clearly shows that already at the initial startup process and proper communication is changing attitude of employees to the issue of abuse of the assets of the employer. Thanks to monitoring of employees is possible to achieve increasing of security in the use of ICT. Such a procedure has a positive impact on the overall safety of the company and all its assets.

Acknowledgment

This publication is supported by the KEGA grant of the Ministry of education, science, research and sport of Slovak Republic No. 021TUKE-4/2012 project (70%) and Slovak Grant Agency of Ministry of Education and Academy of Science of the Slovak Republic within project 1/1147/12 "Methods for analysis of collaborative processes mediated by information systems" (30%).

References

[1] Activity Monitor: [online]. Available on: <http:/www.softactivity.com>
[2] Employee monitoring: When IT is asked to spy: [online]. Available on: <http://www.computerworld.com/s/article/9177981/Employee_monitoring_When_IT_is_asked_to_spy?taxonomyId=17&pageNumber=1>
[3] Employees Steal Data in 5 Ways: [online]. Available on: <http://middleearthbusinesshouse.wordpress.com/tag/employee-monitoring-software/>
[4] Liebert, M., (2004). Development of a Measure of Personal Web Usage in the Workplace CyberPsychology & Behavior, 7, 93-104.
[5] Monitoring Employee Software Comparisons and Reviews 2013: [online]. Available on:<http://employee-monitoring-software-review.toptenreviews.com/>
[6] Monitoring Software Net Spy Pro: [online]. Available on: <http://www.net-monitoring-software.com/>
[7] Monitoring Software Sonar: [online]. Available on: <http://www.awarenesstechnologies.com/sonar.html>
[8] Monitoring Software Spytech: [online]. Available on: <http://www.spytech-web.com/netvizor.shtml>
[9] Monitoring Software Work Examiner: [online]. Available on: <http://www.workexaminer.com/>
[10] Monitorovanie pracovných staníc v komerčnom sektore: [online]. Available on: <http://www.infoconsult.sk/domain/icinfo/files/download/prispevok_sasib_2012.pdf>
[11] Monitorovanie zamestnancov zamestnávateľom [online]. Available on: <http://www.elaw.cz/cs/pracovni-pravo/739-monitorovanie-zamestnancov-zamestnavatelom.html>
[12] MS EAM: [online]. Available on: <http://www.imonitorsoft.com>
[13] MS OsMonitor: [online]. Available on: <http://www.os-monitor.com/>
[14] MS REFOG: [online]. Available on: <http://www.refog.com/>
[15] MS Spector: [online]. Available on: <http://www2.spectorcne.com/>
[16] MS StaffCop: [online]. Available on: <http://www.staffcop.com/>
[17] MS Yaware: [online]. Available on: <http://yaware.com/>
[18] Sociálne siete pod kontrolou zamestnancov [online]. Available on: <http://www.cisco.com/web/SK/news/archive/2010/TS_Socialne_siete_pod_kontrolou_zamestnancov_100413.html>
[19] ŠTOFA, J., a kol.: Process modeling as a supporting tool for managing of the enterprise security, In: IEEE 10th Jubilee International Symposium on Applied Machine Intelligence and Informatics, 26. - 28.1.2012, Herľany, Publisher: Óbuda, Budapest, HU, 2012, 63-67, ISBN 978-1-4577-0195-5), IEEE Catalog Number: CFP1208E-CDR
[20] Viete, čo robia zamestnanci na webe? [online]. Available on: <http://www.itnews.sk/spravy/internet/2012-08-06/c150356-viete-co-robia-zamestnanci-na-webe>
[21] Wakefield, R., (2004). Computer Monitoring and Surveillance: Balancing Privacy with Security. The CPA Journal. 7, 52-55.
[22] Workplace Surveillance and Employee Privacy: Implementing an Effective Computer Use Policy [online]. Available on: <http://www.iima.org/CIIMA/CIIMA%205.2%2057%20Cox-6.pdf>
[23] Záujmy zamestnávateľa vs. súkromie zamestnanca: [online]. Available on: <http://www.isaca.sk/domain/isaca/files/abit-2012/zaujmy-zamestnavatela-vs-sukromie-zamestnanca_judr.-jana-pagacova.pdf>
[24] ZIL, M.: Víte, co dělají vaši zaměstnanci na internetu? In. PC World Security. Roč. 2., č.2 (2005), s.47

Movement Balancing Robot NAO Using Fuzzy Cognitive Maps

Bc. Veronika Szabóová, Supervisor: Dr. Ing. Ján Vaščák
Technical University of Košice,
Letná 9, 04200 Košice, Slovakia
veronika.szaboova@student.tuke.sk, jan.vascak@tuke.sk

***Abstract*—This article examines the balance control of humanoid robot using the pressure sensors on his feet. We present a novel approach where the evolutionary algorithm was used to set up the weights of a fuzzy cognitive map.**

Keywords—fuzzy cognitive map; evolution; stabilization

I. Introduction

According to Kosko [2] most knowledge is specified of classifications and causes. In general, the classes and causes are uncertain (fuzzy or random), usually fuzzy. This fuzziness passes into knowledge representations and on into knowledge bases, where it leads to a knowledge acquisition/ processing trade off'. The fuzzier the knowledge representation, the easier the knowledge acquired and the greater the knowledge-source concurrence. But the fuzzier the knowledge, the harder the (symbolic) knowledge processing.

FCMs are fuzzy-graph structures for representing causal reasoning. Their fuzziness allows vague degrees of causality between vague causal objects (concepts). Their graph structure allows systematic causal propagation, in particular forward and backward chaining, and it allows knowledge bases to be grown by connecting different FCMs. FCMs are especially applicable in soft knowledge domains (e.g. political science, military science, history, international relations, organization theory) where both the system concepts/ relationships and the meta-system language is fundamentally fuzzy.

According to Papageorgiou [5] a method based on evolutionary computation techniques is presented in this paper for training Fuzzy Cognitive Maps. Fuzzy Cognitive Maps is a soft computing technique for modeling complex systems, which combines theories of neural networks and fuzzy logic. The methodology of developing Fuzzy Cognitive Maps relies on human expert experience and knowledge, but still exhibits weaknesses in utilization of learning methods and algorithmic background.

For this purpose, we investigate a coupling of evolutionary algorithm. The use of an evolutionary algorithm is related to the concept of evolution of a number of individuals from generation to generation. The algorithm is introduced, presented and applied in real-world problems as stabilization of a robot. Experimental results suggest that the hybrid strategy is capable to train FCM effectively it leads the system to desired states and determines an appropriate weight matrix for each specific problem.

FCMs are especially applicable to soft knowledge domains and several example FCMs are given. Causality is represented as a fuzzy relation on causal concepts. A fuzzy causal algebra for governing the causal propagation on FCMS is developed. FCM graph representation and matrix operations are presented in the next chapter.

FCM is a hybrid method that lies somewhere between fuzzy systems and neural networks. So FCM represents knowledge by means of symbolic methods and the state space, processes, controls, events, values and inputs in analog methods. FCMS is appropriate to express the knowledge and experience that have been collected over the years for the operation of a complex system. FCM is successfully applied in many areas.[5]

According to Kvasnička [1] modern science is looking for new ideas in the world. So for example, a metaphor of the human brain has led to the emergence of a new area of computer science called neural network. Similarly, the metaphor of Darwinian evolution has led to the emergence of the so-called evolutionary algorithms (EAs), where EA is the most important representative during this last period. The metaphor of Darwinian evolution in computing provides an efficient optimization algorithm that is able to solve difficult tasks. Evolutionary algorithms enable computer simulations of evolutionary processes in biology. These calculations allow to test hypotheses about the meaning and mechanisms of evolution in biology (as well as in other sciences) in the development of various properties [1].

II. Description of fuzzy cognitive maps

According to Kosko [2] the cognitive map (CM) is an oriented weighted graph, where nodes represent concepts and links causal links between them. Each node is assigned a value that indicates whether the concept applies at the moment or not (1 resp. 0). Evaluation of a causal connection between the

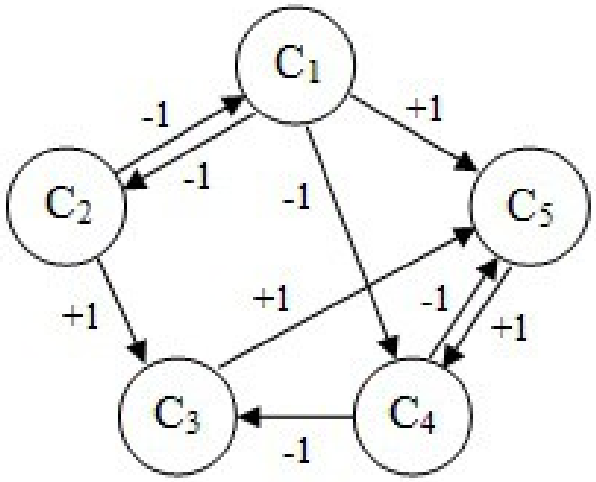


Fig. 1 Example of cognitive map [2]

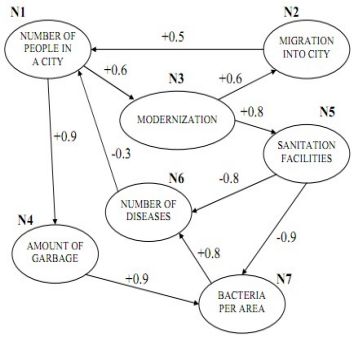
notions of dependence (which may be positive, negative, neutral).

CM is a means for description of complex dynamic systems. The advantage is that this description can be represented graphically and is close to human perception and imagination [2].

According to Stach [3] FCM is an extension of classical CM. The most significant enhancement lies in the way of reflecting causal relationships. Instead of using only the sign (+ or -), each edge is associated with a number (weight) that determines the degree of a considered causal relation between the two concepts. This, in turn, allows implementing knowledge concerning the strength of such a relationship, which now can be described by fuzzy terms, such as weak, medium, strong, or very strong. In other words, a weight of the oriented edge from the node A to B quantifies how much the concept A causes B [3].

The strength of the relationship between two nodes (i.e., weight) is usually normalized to the interval [–1, 1]. The value of –1 represents full negative (inhibitory), +1 full positive, and 0 denotes no causal effect. As a result, an FCM model is fully described by a set of nodes (concepts) and edges (cause-effect relationships), represented by weights, among them. Apart from the graphical representation, for computational purposes, the dependencies captured by this model can be equivalently expressed by a square matrix, called connection matrix, which

stores all weight values for edges between corresponding concepts represented by corresponding rows and columns. An example of the FCM along with its connection matrix is shown in Fig. 2.[3]



	N1	N2	N3	N4	N5	N6	N7
N1	0	0	0.6	0.9	0	0	0
N2	0.5	0	0	0	0	0	0
N3	0	0.6	0	0	0.8	0	0
N4	0	0	0	0	0	0	0.9
N5	0	0	0	0	0	-0.8	-0.9
N6	-0.3	0	0	0	0	0	0
N7	0	0	0	0	0	0.8	0

Fig. 2 FCM model describing public city health issues and its corresponding connection matrix.[3]

The formal definitoujkion of FCM and its markings are following:

Let R be a set of real numbers, N denotes a set of integers and L = [0,1]. Fuzzy cognitive map F is 4-tuple (N, E, C, f), where

1. N = N $\{N_1, N_2, ..., N_n\}$ is the set of n concepts forming the nodes of a graph.

2. E: $(N_i, N_j) \rightarrow e_{ji}$ is a function of N x N to K associating e_{ji} to a pair of concepts (N_i, N_j), with e_{ji} denoting a weight of the directed edge from N_i to N_j if $i \neq j$ and e_{ji} equal zero if i = j. Thus E : (N x N) = $(e_{ji}) \in$ Knxn is a connection matrix.

3. C: $N_i \rightarrow C_i$ is a function that at each concept N_i associates the sequence of its activation degrees such as for $t \in N$, $C_i(t) \in L$ given its activation degree at the moment t. $C(0) \in$ Ln indicates the initial vector and specifies the initial values of all concept nodes and $C(t) \in$ Ln is a state vector at certain iteration t.

4. f : R → L is a transformation function, which includes a recurring relationship on $t \geq 0$ between C(t+1) and C(t)

$$\forall i \epsilon (1,....,n\}, C_i(t+1) = f\left(\sum_{\substack{j=1 \\ i \neq j}}^{n} e_{ji} C_j(t)\right) \quad (1)$$

(1) describes a functional model of FCM, which is used to perform simulations of the system dynamics. Simulation consists of computing the state of the system, which is described by a state vector. The state vector specifies current values of all concepts (nodes) in a particular iteration. The value of a given node is calculated from the preceding iteration values of nodes, which exert influence on the given node through cause-effect relationship (nodes that are connected to the given node) [3].

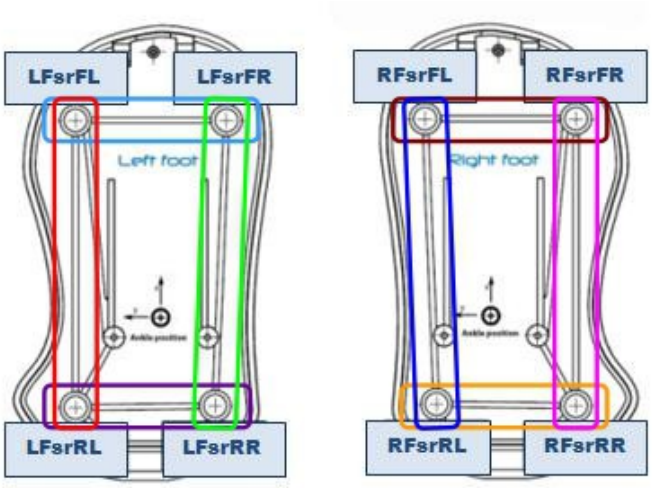


Fig. 3 FCM scheme used to color-coded nodes, depending on the diameter of which two FSR sensors to enter them.

FCM has the capability to incorporate feedback processes. It can be used to simulate the changes of a system over time and address what if questions. In regards of modelling FCM combines aspects of qualitative methods with the advantages of quantitative methods. FCM allows to dynamically simulate the resulting system and to test the influence of management scenarios on system components.

The goal of FCM analysis is to detect and interpreting relations between entities found on a map and to understand its structural properties and dynamics. The structured way of collecting and coding data enables a comparison between different case studies and even aggregation of data. This also fosters the linking of FCM with other management or modeling approaches.

III. Proposal of a system for learning FCM using EA

As basic movements we used a simple walking. We have divided it into 100 sequences, one sequence represents one moment in movement - one pose.

FCM consists of only input and output nodes (Fig. 3).

The number of input nodes is 11 - average of two adjacent embedded FSR sensors and accelerometers X, Y and Z but 100 times because 100 poses are one motion. FSR sensors are the pressure sensors on the bottom of the robots foots.

Nodes of the fuzzy cognitive map contain this data and we teach weights of FCM with the evolutionary algorithm.

An individual is in program represented as a dictionary (python dict) whose keys are "gene" and "fitness" (as example: arg_jedinec={'gen':[-0.5882267406307462, -0.08293475678816042 4,0.21426736876420982, -0.388556...., 0.093987230157366541], 'fitnes': 317.5277242373441}). Key "gene" includes a vector representing the values of all the genes of an individual (values are in the range <-1, 1>). Individual values of genes represent the weights between nodes. Key "fitness" includes an individual fitness evaluation function. The number of genes is (number of input nodes * number of output nodes), which in our case is (11 * 2) (Figure 1). 22 values are therefore used as weights between input and output layer FCM.

3.1. Description of the block algorithm

• Selection – As the first step the elitism is used, i.e. one individual with the best fitness function value is automatically shifted to the next generation. In the second step, two randomly selected individuals that continue to cross (the only requirement is that they must be two different individuals) - the step doubling the number of individuals to cross the block originated from two individuals (parents) only one (offspring).

• Crossing - Each pair of subjects was divided at a random point of the individual and their parts with each other and exchange (single crossover). Output of individuals (offspring) is the first part of the first and second part of the second individual (parent). Crossing does not always perform as described above due to the use of "crossing probability" (in our case 80%) - in this case, the randomly selected one individual of a pair, which is the output of the block intersections.

• Mutation - progress on the "likelihood of mutation" (in this case 5%).

The 5% of cases, the mutation effected by randomly chosen one gene and the value is replaced by a random number generated in the range <-1, 1>. Otherwise, the mutation is not carried out and the algorithm to other blocks in the individual goes through unchanged.

The calculation of the fitness function:

It consists of two steps

• The difference between the desired value and the actual value of the FSR sensor (compared to two as shown in Figure 3). Differences are computed during the execution of the motion (in our case it is 10 seconds of motion) and the amount continues to the next step of the evaluation of the individual.

• Interactive Evolution: carried out on the basis of human users - user evaluates stabilization a robot applied from interval <1, 5>, where 1 is "best" and 5 "worst". This value using the formula ((score - 1) * 100) is added to the fitness function value of the first step.

Interactive evolution is used because it sometimes happens that an individual who had a better course of stabilization, got worse fitness evaluation function (the higher number) as an individual who has been unstable during stabilization.

VI. Conclusions

Fuzzy cognitive maps are generally used for most of the dynamical systems management. However, we have not seen any scientific experiments that FCM is also used to control of balance of a dynamic system. We used FCM for stability when moving. We are now in the process of simulation experimentation and tuning the last elements to be not

damaged by direct robot starts learning stage and life on it. We want this program to test on a real robot trying different speeds, the size of steps and different directions of external interventions on the robot.

References

[1]V.. Kvasnička, Genetic algorithm, Lecture Proceedings, Slovak Technical University, Bratislava, 1998

[2]B. Kosko. Fuzzy cognitive maps. Int. J. Man–Machine Studies, Vol. 24, pp. 65–75, 1986

[3]W. Stach, L. Kurgan, W. Pedrycz and M. Reformat, Evolutionary Development of Fuzzy Cognitive Maps, Fuzzy Systems, FUZZ '05. The 14th IEEE International Conference on, pp. 617-619, 2005

[4]V. Szabóová, Interactive evolution of aesthetic quality of the robot, Bachelor thesis, Technical University of Kosice, 2011

[5]E. I. Papagorgiou, P. P. Groumpos, A new hybrid method using evolutionary algorithms to train Fuzzy Cognitive Maps, Applied soft computing 5 (2005) 1-10

On the Kronecker Product of the Cycles

[1]*Matúš VALO (2^{nd} year)*
Supervisor: [2]Marián KLEŠČ

[1,2]Dept. of Mathematics and Theoretical Informatics, FEI TU of Košice, Slovak Republic

[1]matus.valo@tuke.sk, [2]marian.klesc@tuke.sk

***Abstract*—The crossing number $\mathrm{cr}(G)$ of the graph G is the minimum number of edge crossings in any drawing of G. Regarding the crossing number, the Cartesian product of two graphs has received the greatest attention among the products of graphs. However, some interesting results have appeared recently. This paper presents a survey of known results about the crossing number of the Kronecker product of two graphs and adds an upper bound of the crossing number of the Kronecker product of two cycles.**

***Keywords*—Crossing number, Kronecker product, Cycle**

I. Introduction

A *finite, undirected, simple graph* $G = (V, E)$ is an ordered pair consisting of a finite set V of vertices and a finite set of edges $E \subseteq \{\{u, v\} | u \in V, v \in V, u \neq v\}$. The vertex and edge set of the graph G will be denoted by $V(G)$ and $E(G)$, respectively. We will abuse terminology and therefore a finite, undirected, simple graph will be called a *graph*. The graph is also connected unless it is not explicitly stated. An edge will be denoted by uv instead of $\{u, v\}$. A graph $G = (V, E)$ is called *bipartite* if V admits a partition into two classes such that every edge has its endvertices in different classes. A path and a cycle with n vertices will be denoted by P_n and C_n, respectively.

Graphs are usually represented in the plane by visualising each vertex as a single point and each edge as a curve connecting points which represent end vertices of the given edge. We will consider drawings such that any two edges:

- intersect in a finite number of points,
- do not touch and
- for the crossing x there are at most two edges of G whose crossing is x.

The *crossing number* of a graph G is denoted by $\mathrm{cr}(G)$ and is defined as the smallest number of edge crossings in any drawing of G. A drawing with minimum number of crossings (an optimal drawing) must be a *good* drawing, that is a drawing satisfying the following properties:

- no edge crosses itself,
- no two edges cross more than once and
- no two adjacent edges cross.

Finding a crossing number is useful in many areas. The most prominent area is VLSI technology. The lower bound on the chip area is determined by crossing number and by number of vertices of the graph [3], [14]. It plays an important role in various fields of discrete/computational geometry [16]. The crossing number is also a parameter yielding the deviation of the graph from being planar. The crossing number significantly influences readability and therefore it is the most important parameter when considering aesthetics of a graph. It is mostly used in automated visualisation of graphs.

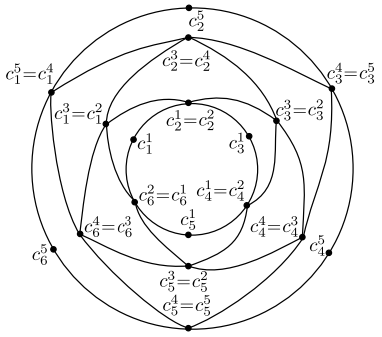


Fig. 1. A drawing of the graph $C_{3,5}$.

However, finding a crossing number of a given graph is very hard. It was proved by Garey and Johnson [7] that the crossing number of a graph is NP–complete and it remains NP–complete even for cubic graphs [9].

A very special type of drawing of a graph is an orthogonal drawing, which allows only $\pi/2$ or π angle between two adjacent edges. This type of drawing has several applications.

The basic abstraction underlying a software system running on a set of distinct machines consists of a set of finite transition systems. An orthogonal drawing gives an intuitive representation of such systems [6].

Computer hardware and microchips are designed using CAD tools, which must create a layout of the logic gates and their interconnections on circuit boards. The layout itself corresponds to a grid drawing in which all vertices and bends of the edges have integer coordinates [2]. An orthogonal drawing is similar to a grid drawing.

Entity–relationship diagrams are one of the common methods to aid structuring large volumes of data by defining attributes and relations between the data [6]. Both entities and their attributes are modeled as nodes of a graph. Edges express that the given attribute is an attribute of the given entity by linking an attribute to an entity. The resulting graph is presented by means of an orthogonal drawing.

II. The Kronecker Product

Let G and H be graphs. The *Kronecker product* $G \times H$ is a graph with the vertex set $V(G \times H) = V(G) \times V(H)$ and the edge set $E(G \times H) = \{\{(x_g, x_h), (y_g, y_h)\} | x_g y_g \in E(G)$ and $x_h y_h \in E(H)\}$. This product is also known

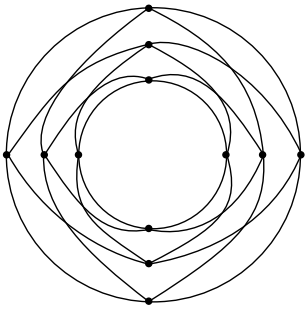


Fig. 2. A drawing of the graph $C_3 \times C_4$.

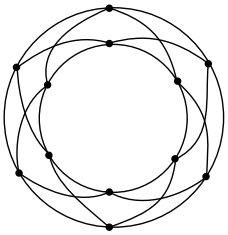


Fig. 3. A drawing of the graph $C_3 \times C_4$.

as a direct product, tensor product, cardinal product, cross product and graph conjunction.

The Kronecker product has plenty interesting properties. P. M. Weichsel proved assertion that $G \times H$ is connected if and only if either G or H contains an odd cycle. If neither of them contains an odd cycle then $G \times H$ contains exactly two connected components [19].

A bipartite graph $G = (V_0 \cup V_1, E)$ is said to have a property π if G admits an automorphism ϕ such that $x \in V_0$ if and only if $\phi(x) \in V_1$. Jha, Klavžar and Zmazek showed the following [13]:

Theorem 1: If G and H are bipartite graphs one of which has property π, then two components of $G \times H$ are isomorphic.

Jha, Klavžar and Zmazek also conjectured that the converse of the Theorem holds. This conjecture was later proved by Hammack [8]:

Theorem 2: Suppose G and H are connected bipartite graphs. The two components of $G \times H$ are isomorphic if and only if at least one of G or H has property π.

Let a 1-contraction be the removal each vertex of degree 1 from G. Farzan and Waller proved the following Theorem [5]:

Theorem 3: Let G_1 and G_2 be connected graphs with more than four vertices. Then $G_1 \times G_2$ is planar if and only if either

1) one of the graphs is a path and the other one is 1-contractible to a path or circuit, or
2) one of them is a circuit and the other is 1-contractible to a path.

III. The Crossing Number of $C_m \times C_n$.

The Kronecker product $C_m \times C_n$ is an important class, which has many useful properties. The most prominent property is that such a graph is a four–regular graph, furthermore, it contains subgraphs isomorphic to a grid. It enables an orthogonal drawing of such graph. Jha and Devisetty formulated algorithms for orthogonal embedding $C_m \times C_n$ [12].

The *Cartesian product* $G \square H$ of the graphs G and H is a graph with the vertex set $V(G \square H) = V(G) \times V(H)$ and the edge set $E(G \square H) = \{\{(x_g, x_h), (y_g, y_h)\} | x_g y_g \in E(G)$ and $x_h = y_h,$ or $x_h y_h \in E(H)$ and $x_g = y_g\}$. If m and n are both odd, then $C_m \times C_n$, also known as diagonal mesh [17], [11], [18], has a lower diameter, higher independence number and higher odd girth relative to $C_m \square C_n$, also known as a toroidal mesh. Pearlmutter showed that a diagonal mesh is isomorphic to a twisted toroidal mesh [15] and that a twisted toroidal topology was earlier used as the routing network of the FAIM–1 parallel computer [1].

The crossing number of $C_m \times C_n$ is so far unknown, but Jha and Devisetty proved its lower bound [12]:

Theorem 4: $\mathrm{cr}(C_m \times C_n)$ is greater than or equal to

1)	$(0.8 - \epsilon)mn$	m, n is odd and equal, $m \geq n_0$
2)	$\frac{1}{8}(0.8 - \epsilon)mn$	m is odd, n is even, $n \geq 6$, $\min\{m, \frac{n}{2}\} \geq n_0$
3)	$\frac{1}{8}(0.8 - \epsilon)m(n-1)$	m, n is odd, $m < n, n \geq 7$, $\min\{m, \frac{(n-1)}{2}\} \geq n_0$

where, in each case, $\epsilon > 0$ and n_0 is a sufficiently large integer depending only on ϵ.

The proved upper bound too:

Theorem 5: Let m be odd. Then $\mathrm{cr}(C_m \times C_n)$ is less than or equal to

1)	$(m-2)n$	$n = m$
2)	$(m-1)n$	$n = km, k$ is even, $k \geq 2$
3)	$(m-1)n + m$	$n = km, k$ is odd, $k \geq 3$
4)	$(2m-n)(n-3) + n(\frac{n}{2} - 1)$	n is even, $m \geq \frac{n}{2}$
5)	$(n-2m)(2m-3) + 2m(m-1)$	n is even, $\frac{n}{2} > m$
6)	$2mn - (m^2 + 3m)$	n is odd, $m < n < 3m$
7)	$(2mn + 6m) - (2m^2 + 3n)$	n is odd, $n \geq 3m$.

In the rest of this section, we give the upper bound for the crossing number of the Kronecker products of two cycles C_m and C_n for odd $m \geq 3$ and even $n \geq 4$. We do not address the case when m and n are both even, since $C_m \times C_n$ in that case consists of two connected components isomorphic to each other, and it turns out that each such component is similar to $C_m \times C_n$ where m is odd and n is even [13].

Assume $m \geq 2,\ n \geq 1$. It seems convenient to consider a graph $C_{m,n}$ in the following way: it contains n edge disjoint cycles $C^i = \{c^i_1, c^i_2, \ldots, c^i_{2m}\},\ i = 1, \ldots, n-1, n$, of length $2m$ and

- $V(C^i) \cap V(C^{i+1}) = \{c^i_1, c^i_3, \ldots, c^i_{2m-1}\}$ and $V(C^i) \cap V(C^{i-1}) = \{c^i_2, c^i_4, \ldots, c^i_{2m}\}$, if i is even,
- $V(C^i) \cap V(C^{i-1}) = \{c^i_1, c^i_3, \ldots, c^i_{2m-1}\}$ and $V(C^i) \cap V(C^{i+1}) = \{c^i_2, c^i_4, \ldots, c^i_{2m}\}$, if i is odd.

An example of the graph $C_{3,5}$ is shown in Fig. 1. It is said that C^i and $C^j,\ i \neq j$, are adjacent if $V(C^i) \cap V(C^j) \neq \emptyset$. Hence $C^1,\ C^n$ are adjacent to only one cycle $C^2,\ C^{n-1}$, respectively, and $C^i,\ i = 2, 3, \ldots n-1$, is adjacent to either C^{i-1} and C^{i+1}. It can be easily seen that $C_{m,n}$ is isomorphic to $C_m \times P_{n+1}$ if m is odd number or to both components of the graph $C_{2m} \times P_{n+1}$ if m is even number. Moreover, from Fig. 1, it is clear that the following Lemma holds:

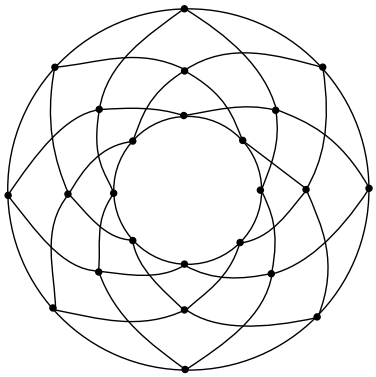


Fig. 4. A drawing of the graph $C_3 \times C_8$.

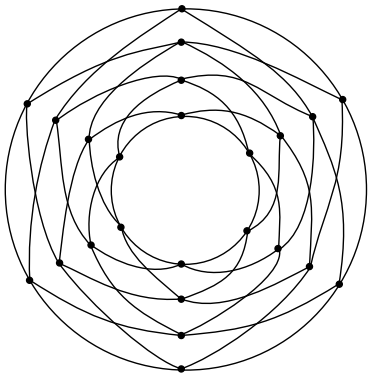


Fig. 5. A drawing of the graph $C_3 \times C_8$.

Lemma 1: $\mathrm{cr}(C_{m,n}) = 0,\ m \geq 2,\ n \geq 1.$

Let $m \geq 3$ be an odd and $n \geq 4$ be an even number. There are two possibilities of constructing the graph $C_m \times C_n$:

- from two copies $C_{m,\frac{n}{2}+1}$, $C'_{m,\frac{n}{2}+1}$ by identifying both cycles C^1 with C'^1, $C^{\frac{n}{2}+1}$ with $C'^{\frac{n}{2}+1}$, respectively, or
- from two copies $C_{\frac{n}{2},m+1}$, $C'_{\frac{n}{2},m+1}$ by identifying both cycles C^1 with C'^1, C^{m+1} with C'^{m+1}, respectively.

Hence a drawing of $C_m \times C_n$ can be constructed from the graphs $C_{m,\frac{n}{2}+1}$, $C'_{m,\frac{n}{2}+1}$ by identifying both cycles C^1 with C'^1, $C^{\frac{n}{2}+1}$ with $C'^{\frac{n}{2}+1}$, respectively, and by placing every vertex $c'^i_j \in V(C'_{m,\frac{n}{2}+1}) \setminus (V(C'^1) \cup V(C'^{\frac{n}{2}+1}))$ into the face of the graph $C_{m,\frac{n}{2}+1}$ induced by the vertices $c^{i-1}_j, c^{i+1}_j, c^i_{j-1}, c^i_{j+1}$. This construction yields the drawing of the graph $C_3 \times C_8$ shown in Fig. 5.

In the similar way, the graph $C_m \times C_n$ can be constructed from $C_{\frac{n}{2},m+1}$, $C'_{\frac{n}{2},m+1}$ by identifying both cycles C^1 with C'^1, C^{m+1} with C'^{m+1}, respectively, and by placing every vertex $c'^i_j \in V(C'_{\frac{n}{2},m+1}) \setminus (V(C'^1) \cup V(C'^{m+1}))$ into the face of the graph $C_{\frac{n}{2},m+1}$ induced by the vertices $c^{i-1}_j, c^{i+1}_j, c^i_{j-1}, c^i_{j+1}$. This construction yields the drawing of the graph $C_3 \times C_8$ shown in Fig. 4.

Given constructions yield different drawings with a different number of crossings. It can be easily seen that the drawing of the graph $C_3 \times C_4$ in Fig. 3 is created from two copies of $C_{3,3}$ yielding six crossings, but the drawing of the same graph in Fig. 2 is created from two copies of $C_{2,4}$ yielding eight crossings. Hence the first construction yields less crossings for the graph $C_3 \times C_4$. But when the graph $C_3 \times C_8$ is considered, the drawing created from two copies of $C_{3,5}$, showed in Fig. 5, yields 18 crossings. On the other hand the drawing from two copies of $C_{4,4}$, showed in Fig. 4, yields only 16 crossings. This implies that the second construction yields less crossings for the graph $C_3 \times C_8$. Therefore, there are two sets of the graphs $C_m \times C_n$ differing in construction yielding a drawing with less crossings.

When the graph $C_m \times C_n$ is constructed by the first method, every cycle C^i is crossed by the cycle C'^i exactly $2m$ times for every $i = 2, 3, \ldots, \frac{n}{2}$. Hence, in this drawing, the number of crossings is equal to $2m(\frac{n}{2} - 1) = m(n-2)$.

Despite that, the second method yields the drawing in such way that every cycle C^i is crossed by the cycle C'^i exactly n times for every $i = 2, 3, \ldots, m$. Hence, in this drawing, the number of crossings is equal to $n(m-1)$.

As, for odd $m \geq 3$ and even $n \geq 4$, we did not find a drawing of the graph $C_m \times C_n$ with less than $m(n-2)$ or $n(m-1)$ crossings, we can formulate the next conjecture.

Conjecture 1: Let m be an odd and n be an even number. Then $\mathrm{cr}(C_m \times C_n) = \min\{m(n-2), n(m-1)\}$, for $m \geq 3,\ n \geq 4$.

References

[1] J. M. Anderson, W. S. Coates, A. L. Davis, R. W. Hon, I. N. Robinson, S. V. Robison, and K. S. Stevens, "The architecture of FAIM-1," *IEEE Computer*, 1987.

[2] G. Di Battista, P. Eades, R. Tamassia and I. G. Tollis, "Graph Drawing: Algorithms for the Visualization of Graphs". *Prentice Hall*, Jul. 1998.

[3] S. N. Bhatt and F. T. Leighton, "A framework for solving VLSI graph layout problems," *Journal of Computer and System Sciences*, vol. 28, pp. 300–343, 1984.

[4] A. Bottreau and Y. Métivier, "Some remarks on the kronecker product of graphs," *Information Processing Letters*, vol. 68, no. 2, pp. 55–61, 1998.

[5] M. Farzan and D. A. Waller, "Kronecker products and local joins of graphs," *Canad. J. Math*, vol. 29, pp. 255–269, 1977.

[6] R. Fleischer and C. Hirsh, "Drawing graphs," M. Kaufmann and D. Wagner, Eds. London, UK: Springer-Verlag, 2001, ch. Graph drawing and its applications, pp. 1–22.

[7] M. R. Garey and D. S. Johnson, "Crossing number is NP-complete," *SIAM Journal on Algebraic and Discrete Methods*, vol. 4, no. 3, pp. 312–316, 1983.

[8] R. H. Hammack, "Proof of a conjecture concerning the direct product of bipartite graphs, " *Journal European Journal of Combinatorics*, vol. 30, pp. 1114–1118, 2009.

[9] P. Hliněný, "Crossing number is hard for cubic graphs," *J. Comb. Theory Ser. B*, vol. 96, pp. 455–471, July 2006.

[10] P. K. Jha, "Kronecker products of paths and cycles: Decomposition, factorization and bi-pancyclicity," *Discrete Mathematics*, vol. 182, no. 1–3, pp. 153–167, 1998.

[11] P. K. Jha, "A counterexample to tang and padubidri's claim about the bisection width of a diagonal mesh," *IEEE Transactions on Computers*, vol. 52, pp. 676–677, 2003.

[12] P. K. Jha and S. Devisetty, "Orthogonal drawings and crossing numbers of the kronecker product of two cycles," *Journal of Parallel and Distributed Computing*, vol. 72, no. 2, pp. 195–204, 2012.

[13] P. K. Jha, S. Klavžar and B. Zmazek, "Isomorphic Components of Kronecker Product of Bipartite Graphs," *Discussiones Mathematicae Graph Theory*, vol. 17, pp. 301–309, 1997.

[14] F. T. Leighton, "Complexity issues in VLSI: optimal layouts for the shuffle-exchange graph and other networks". Cambridge, MA, USA: MIT Press, 1983.

[15] B. A. Pearlmutter, "Doing the twist: Diagonal meshes are isomorphic to twisted toroidal meshes," *IEEE Transactions on Computers*, vol. 45, pp. 766–768, 1996.

[16] F. Shahrokhi, O. Sýkora, L. A. Székely, and I. Vrťo, "Intersection of curves and crossing number of $C_m \times C_n$ on surfaces," *Discrete & Computational Geometry*, vol. 19, pp. 237–247, 1998, 10.1007/PL00009343.

[17] K. W. Tang and S. A. Padubidri, "Diagonal and toroidal mesh networks," *IEEE Trans. Comput*, vol. 43, pp. 815–826, 1994.

[18] K. W. Tang and R. Kamoua, "An upper bound for the bisection width of a diagonal mesh," *IEEE Trans. Comput.*, vol. 56, pp. 429–431, 2007.

[19] P. M. Weichsel, "The kronecker product of graphs," *Proceedings of the American Mathematical Society*, vol. 13, no. 1, pp. 47–52, 1962.

Output controller design with output variables constraint

Vladimír SERBÁK (2^{nd} year), Pavol LIŠČINSKÝ (1^{st} year)
Supervisor: Anna FILASOVÁ

Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

vladimir.serbak@tuke.sk, pavol.liscinsky@tuke.sk

***Abstract*—In this paper, we consider the problem of design of stabilizing control for linear discrete-time systems, reflecting equality constraint tying together chosen output variables. The proposed designs reflects equality constraint tying together chosen output variables in desired ratio. Both algorithms compared in this paper are computationally simple and connected with Lyapunov stability theory. The approach is illustrated on simulation examples, where the validity of the proposed methods are demonstrated.**

***Keywords*—linear matrix inequalities, optimal control, static output feedback, output equality constraint**

I. Introduction

In the last years many results have spurred great interest for many applications in the problem of control design for systems with constraint. In the typical case, where a system state reflects certain constraint defined by physical limits, these constraint usually keep the system state in the required region of technological conditions [1], [2].

In principle, it is possible and ever easy to design a state controller that stabilizes the systems and simultaneously forces closed-loop systems to satisfy constraint, although in case of static output feedback controller given task can be more challenging or even impossible [3], [4]. Various approaches for obtaining the feedback matrix for asymptotic stabilization of linear systems have been developed.

Linear matrix inequalities (LMI's) offer additional approach to solve various tasks in control theory [5]. Various tasks in the static output feedback (SOF) control, even those absent of analytical solution, can be approached by reduction to convex and quasi-convex problems involving matrix inequalities [6]. Such an adaptation has a useful practical interest due to the presence of powerful numerical solvers [7].

Optimal control deals with the problem of finding a control law for a given system such that a certain optimality criterion (performance index) is achieved. The system which is the end result of an optimal design is not supposed merely to satisfy some of the constrains associated with the classical control (stability, bandwidth ...) but it is supposed to be the best possible system of a particular type. In optimal control problems, variables are separated into two classes, namely the state variables and the control variables. In proposed approach those classes are transformed to the output variables and to the modified control variables. Further, the output as well as the modified control variables is generally subject to constraint, which make many problems in optimal control non-classical, since problems with path constraint can hardly be handled in the classical calculus of variations. The linear quadratic regulator (LQR) is a special case of optimal control problem that can be analytically solved, which arises when the objective (cost) function is a quadratic function of state q and input u, and a linear discrete-time state equation is considered [8], [9].

The problem of SOF controller design is studied in this paper. Both proposed design approaches allows to include equality constraint for discrete time systems tying together chosen output variables. The considered task in the paper is to determine output feedback control of discrete-time linear systems forcing output variables of a linear system to satisfy demanded constraint and achieve asymptotic stability and compare the results. The structure of generalized controllers are based on state description of discrete-time linear system. Proposed type of control is suitable for ratio control of output variables.

The paper is organized as follows. Section II defines formulation of the given problem. Section III communicates basic preliminaries for further sections. Section IV contains detailed description of constrained output control design for both LMI and optimal control approach. Section V contains illustrative example of the proposed approaches.

II. Problem formulation

The task is for discrete-time linear multi-input/multi-output dynamic system

$$\boldsymbol{q}(i+1) = \boldsymbol{F}\boldsymbol{q}(i) + \boldsymbol{G}\boldsymbol{u}(i) \tag{1}$$

$$\boldsymbol{y}(i) = \boldsymbol{C}\boldsymbol{q}(i) \tag{2}$$

to design an asymptotically stable closed loop system using an output feedback controller of the form

$$\boldsymbol{u}(i) = -\boldsymbol{K}\boldsymbol{y}(i) = -\boldsymbol{K}\boldsymbol{C}\boldsymbol{q}(i) \tag{3}$$

where $\boldsymbol{q}(i) \in I\!R^n, \boldsymbol{u}(i) \in I\!R^r, \boldsymbol{y}(i) \in I\!R^m$ are vectors of the state, input and output variables, respectively and nominal system matrices $\boldsymbol{F} \in I\!R^{n\times n}, \boldsymbol{G} \in I\!R^{n\times r}$ and $\boldsymbol{C} \in I\!R^{m\times n}$ are real matrices.

While all output variables are measurable, $\boldsymbol{K} \in I\!R^{r\times m}$ is the feedback controller gain matrix, and design constraint in the next equality form

$$\boldsymbol{D}\boldsymbol{y}(i) = 0 \tag{4}$$

is considered with $\boldsymbol{D} \in I\!R^{k\times m}$, rank $\boldsymbol{D} = k \leq r$.

In case of the optimal control design, the task is to determine the output control able to minimize the quadratic cost function

$$\boldsymbol{J_N} = \boldsymbol{y}^T(N)\boldsymbol{Q}^\circ\boldsymbol{y}(N) + \sum_{i=0}^{N-1} s(\boldsymbol{y}(i), \boldsymbol{u}(i)) \tag{5}$$

$$s(\boldsymbol{y}(i), \boldsymbol{u}(i)) = [\boldsymbol{y^T}(i)\ \boldsymbol{u^T}(i)] \begin{bmatrix} \boldsymbol{Q} & \boldsymbol{S} \\ \boldsymbol{S}^T & \boldsymbol{R} \end{bmatrix} \begin{bmatrix} \boldsymbol{y}(i) \\ \boldsymbol{u}(i) \end{bmatrix} \tag{6}$$

$$\boldsymbol{Q} - \boldsymbol{S}\boldsymbol{R}^{-1}\boldsymbol{S}^T \geq 0 \tag{7}$$

where matrices $\boldsymbol{Q} \geq 0 \in I\!R^{m\times m}$, $\boldsymbol{Q}^\circ \geq 0 \in I\!R^{m\times m}$ and $\boldsymbol{R} \geq 0 \in I\!R^{r\times r}$ has full row rank, $\boldsymbol{S} \in I\!R^{m\times r}$ satisfies (7), $\boldsymbol{K} \in I\!R^{r\times m}$ is the gain matrix of the feedback controller.

III. Basic preliminaries

Proposition 1: Let $\boldsymbol{X}$ be a matrix variable and $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{Y}$ are known non-square matrices of appropriate dimensions such the equality

$$\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X}\boldsymbol{B} \tag{8}$$

can be set then all solution to $\boldsymbol{X}$ means

$$\boldsymbol{X} = \boldsymbol{A}^{\ominus 1}\boldsymbol{Y}\boldsymbol{B}^{\ominus 1} + \boldsymbol{Z} - \boldsymbol{A}^{\ominus 1}\boldsymbol{A}\boldsymbol{Z}\boldsymbol{B}\boldsymbol{B}^{\ominus 1} \tag{9}$$

where

$$\boldsymbol{B}^{\ominus 1} = (\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{B}^T \tag{10}$$

$$\boldsymbol{A}^{\ominus 1} = \boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{A}^T)^{-1} \tag{11}$$

are Moore - Penrose pseudoinverses of $\boldsymbol{B}$ and $\boldsymbol{A}$, and $\boldsymbol{Z}$ is an arbitrary matrix of appropriate dimension [2].

Proposition 2: (Schur complement)

Let $\boldsymbol{Q} > 0, \boldsymbol{R} > 0, \boldsymbol{S}$ are real matrices of appropriate dimensions, then the next inequalities are equivalent [8]

$$\begin{bmatrix} \boldsymbol{Q} & \boldsymbol{S} \\ \boldsymbol{S}^T & \boldsymbol{R} \end{bmatrix} > 0 \Leftrightarrow \begin{bmatrix} \boldsymbol{Q} - \boldsymbol{S}\boldsymbol{R}^{-1}\boldsymbol{S}^T & 0 \\ 0 & \boldsymbol{R} \end{bmatrix} > 0 \tag{12}$$

$$\boldsymbol{Q} - \boldsymbol{S}\boldsymbol{R}^{-1}\boldsymbol{S}^T > 0, \boldsymbol{R} > 0 \tag{13}$$

IV. Constrained output control design

The format of the stabilization problem with the pure matrix algebraic equation constraint is prescribed by a matrix $\boldsymbol{D} \in I\!R^{k\times m}$ to give the design constraint (4) implying that the output variable vectors have to satisfy equalities

$$\boldsymbol{D}\boldsymbol{C}\boldsymbol{q}(i+1) = \boldsymbol{D}\boldsymbol{C}(\boldsymbol{F} - \boldsymbol{G}\boldsymbol{K}\boldsymbol{C})\boldsymbol{q}(i) = 0 \tag{14}$$

for $i = 1, 2, \ldots$.

It is supposed the matrix D is chosen by such way that

$$\boldsymbol{D}\boldsymbol{C}(\boldsymbol{F} - \boldsymbol{G}\boldsymbol{K}\boldsymbol{C}) = 0 \tag{15}$$

$$\boldsymbol{D}\boldsymbol{C}\boldsymbol{F} = \boldsymbol{D}\boldsymbol{C}\boldsymbol{G}\boldsymbol{K}\boldsymbol{C} \tag{16}$$

respectively, as well as that the closed-loop system matrix $(\boldsymbol{F} - \boldsymbol{G}\boldsymbol{K}\boldsymbol{C})$ is stable. Solving (15) with respect to $\boldsymbol{K}$ using (9) all solutions of $\boldsymbol{K}$ as follows

$$\begin{aligned} \boldsymbol{K} = (\boldsymbol{D}\boldsymbol{C}\boldsymbol{G})^{\ominus 1}\boldsymbol{D}\boldsymbol{C}\boldsymbol{F}\boldsymbol{C}^{\ominus 1} + \boldsymbol{J} - \\ -(\boldsymbol{D}\boldsymbol{C}\boldsymbol{G})^{\ominus 1}(\boldsymbol{D}\boldsymbol{C}\boldsymbol{G})\boldsymbol{J}\boldsymbol{C}\boldsymbol{C}^{\ominus 1} \end{aligned} \tag{17}$$

where J is an arbitrary matrix with appropriate dimensions. thus it is possible express $\boldsymbol{K}$ as

$$\boldsymbol{K} = \boldsymbol{M} + \boldsymbol{N}\boldsymbol{J}^* \tag{18}$$

where $\boldsymbol{M} \in I\!R^{r\times m}$, $\boldsymbol{N} \in I\!R^{r\times 2r}$ and $\boldsymbol{J}^* \in I\!R^{2r\times m}$

$$\boldsymbol{M} = (\boldsymbol{D}\boldsymbol{C}\boldsymbol{G})^{\ominus 1}\boldsymbol{D}\boldsymbol{C}\boldsymbol{F}\boldsymbol{C}^{\ominus 1} \tag{19}$$

$$\boldsymbol{N} = \begin{bmatrix} \boldsymbol{I} & -(\boldsymbol{D}\boldsymbol{C}\boldsymbol{G})^{\ominus 1}\boldsymbol{D}\boldsymbol{C}\boldsymbol{G} \end{bmatrix} \tag{20}$$

$$\boldsymbol{J}^* = \begin{bmatrix} \boldsymbol{J} \\ \boldsymbol{J}\boldsymbol{C}\boldsymbol{C}^{\ominus 1} \end{bmatrix} \tag{21}$$

A. LMI approach

Theorem 1: The system (1), (2) controlled by the output controller (3), (18) with design constraint (4), is global asymptotically stable if there exists a positive definite symmetric matrix $\boldsymbol{W} \in I\!R^{n\times n}$ and matrices $\boldsymbol{T} \in I\!R^{m\times m}$, $\boldsymbol{Z} \in I\!R^{2r\times n}$ such that

$$\boldsymbol{W} = \boldsymbol{W}^T > 0 \tag{22}$$

$$\begin{bmatrix} -\boldsymbol{W} & \boldsymbol{W}\boldsymbol{F}_1^T - \boldsymbol{Z}^T\boldsymbol{G}_1^T \\ * & -\boldsymbol{W} \end{bmatrix} < 0 \tag{23}$$

$$\begin{bmatrix} -\varepsilon\boldsymbol{I} & \boldsymbol{C}\boldsymbol{W} - \boldsymbol{T}\boldsymbol{C} \\ * & -\boldsymbol{I} \end{bmatrix} < 0 \tag{24}$$

If above conditions (22), (23), (24) hold, the control law gain matrix $\boldsymbol{J}^* \in I\!R^{r\times m}$ is given by

$$\boldsymbol{J}^* = \boldsymbol{Z}(\boldsymbol{T}\boldsymbol{C})^{\ominus 1} = \boldsymbol{Z}(\boldsymbol{G}\boldsymbol{W})^{\ominus 1} \tag{25}$$

and the feedback controller gain matrix $\boldsymbol{K} \in I\!R^{r\times m}$ is given by (18) where matrices $\boldsymbol{M} \in I\!R^{r\times m}$ and $\boldsymbol{N} \in I\!R^{r\times 2r}$ are given by (19), (20).

Proof: See [11]. ■

B. Optimal control approach

Theorem 2: The system (1), (2) controlled by the output controller

$$\boldsymbol{u}(i) = -\boldsymbol{K}\boldsymbol{y}(i) = -\boldsymbol{M}\boldsymbol{y}(i) + \boldsymbol{N}\widetilde{\boldsymbol{u}}(i) \tag{26}$$

$$\widetilde{\boldsymbol{u}}(i) = -\boldsymbol{J}^*\boldsymbol{y}(i) \tag{27}$$

with design constraint (4), is global asymptotically stable if there exists a positive definite symmetric matrix $\boldsymbol{P} = \boldsymbol{P}^T > 0$ of discrete Riccati equation in the form

$$\begin{aligned} \boldsymbol{E}^T\boldsymbol{P}(i-1)\boldsymbol{E} = \boldsymbol{F}^{\circ T}\boldsymbol{P}(i)\boldsymbol{F}^\circ + \boldsymbol{Q}^\circ - (\boldsymbol{F}^{\circ T}\boldsymbol{P}(i)\boldsymbol{G}^\circ + \boldsymbol{S}^\circ). \\ .(\boldsymbol{G}^{\circ T}\boldsymbol{P}(i)\boldsymbol{G}^\circ + \boldsymbol{R}^\circ)^{-1}(\boldsymbol{F}^{\circ T}\boldsymbol{P}(i)\boldsymbol{G}^\circ + \boldsymbol{S}^\circ)^T \end{aligned} \tag{28}$$

Optimal control constraint in output by equality constraint is given in form (26) where

$$\widetilde{\boldsymbol{u}}(i) = -[\boldsymbol{G}^{\circ T}\boldsymbol{P}(i)\boldsymbol{G}^\circ + \boldsymbol{R}^\circ]^{-1}[\boldsymbol{F}^{\circ T}\boldsymbol{P}(i)\boldsymbol{G}^\circ + \boldsymbol{S}^\circ]^T\boldsymbol{y}(i) \tag{29}$$

$$\begin{aligned} \boldsymbol{E} &= \boldsymbol{C}\boldsymbol{C}^{\ominus 1} \\ \boldsymbol{F}^\circ &= \boldsymbol{C}\boldsymbol{F}\boldsymbol{C}^{\ominus 1} - \boldsymbol{C}\boldsymbol{G}\boldsymbol{M} \\ \boldsymbol{G}^\circ &= \boldsymbol{C}\boldsymbol{G}\boldsymbol{N} \\ \boldsymbol{S}^\circ &= (\boldsymbol{C}^{\ominus 1})^T\boldsymbol{C}^T\boldsymbol{S}\boldsymbol{N} - \boldsymbol{M}^T\boldsymbol{R}\boldsymbol{N} \\ \boldsymbol{R}^\circ &= \boldsymbol{N}^T\boldsymbol{R}\boldsymbol{N} \\ \boldsymbol{Q}^\circ &= (\boldsymbol{C}^{\ominus 1})^T\boldsymbol{C}^T\boldsymbol{Q}\boldsymbol{C}\boldsymbol{C}^{\ominus 1} - (\boldsymbol{C}^{\ominus 1})^T\boldsymbol{C}^T\boldsymbol{S}\boldsymbol{M} - \\ &\quad -\boldsymbol{M}^T(\boldsymbol{C}^T\boldsymbol{S})^T\boldsymbol{C}^{\ominus 1} + \boldsymbol{M}^T\boldsymbol{R}\boldsymbol{M} \end{aligned} \tag{30}$$

and $\boldsymbol{E} \in I\!R^{m\times m}$, $\boldsymbol{F}^\circ \in I\!R^{m\times m}$, $\boldsymbol{G}^\circ \in I\!R^{r\times 2r}$, $\boldsymbol{S}^\circ \in I\!R^{m\times 2r}$, $\boldsymbol{R}^\circ \in I\!R^{2r\times 2r}$, $\boldsymbol{Q}^\circ \in I\!R^{m\times m}$ and matrices $\boldsymbol{M}$ and $\boldsymbol{N}$ are given by (19) and (20).

Proof: See [12]. ■

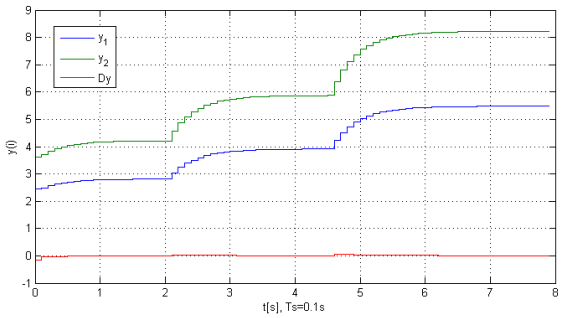


Fig. 1. Response of the system in forced mode with output controller designed by LMI approach.

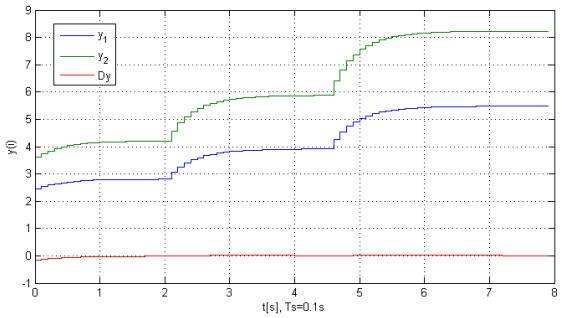


Fig. 2. Response of the system in forced mode with output controller designed by optimal control approach.

V. Constrained forced mode

The output control in unforced mode was defined by

$$\boldsymbol{u}(i) = -\boldsymbol{K}\boldsymbol{y}(i) = -\boldsymbol{K}\boldsymbol{C}\boldsymbol{q}(i) \tag{31}$$

and the system outputs were constrained in this subspace (the null space of $\boldsymbol{D}$) and stay within the constraint subspace, i.e. $(\boldsymbol{F} - \boldsymbol{G}\boldsymbol{K}\boldsymbol{C})\boldsymbol{q}(i) \in N_D$ [9].

The output control in a forced mode is defined by

$$\boldsymbol{u}(i) = -\boldsymbol{K}\boldsymbol{y}(i) + \boldsymbol{W}_w\boldsymbol{w}(i) \tag{32}$$

where $\boldsymbol{w}(i) \in I\!R^m$ is a desired output vector signal, and $\boldsymbol{W}_w \in I\!R^{r\times m}$ is the signal gain matrix.

Matrix $\boldsymbol{W}_w$ can be designed as

$$\boldsymbol{W}_w = (\boldsymbol{C}(\boldsymbol{I}_m - (\boldsymbol{F} - \boldsymbol{G}\boldsymbol{K}))^{-1}\boldsymbol{G})^{-1} \tag{33}$$

for the equations (1) and (2) if the following condition is satisfied

$$rank\begin{bmatrix}\boldsymbol{F} & \boldsymbol{G}\\ \boldsymbol{C} & \boldsymbol{D}\end{bmatrix} = \boldsymbol{n} + \boldsymbol{r} \tag{34}$$

Theorem 3: If the closed-loop system state variables satisfy the consraint (4), then the common output variable vector $\boldsymbol{y}_d(i) = \boldsymbol{D}\boldsymbol{y}(i), \boldsymbol{y}_d(i) \in I\!R^k$ attains the steady-state value

$$\boldsymbol{y}_d = \boldsymbol{D}\boldsymbol{C}\boldsymbol{G}\boldsymbol{W}_w\boldsymbol{w}_s \tag{35}$$

where $\boldsymbol{w}_s$ is the desired signal and $\boldsymbol{W}_w$ is given by (33).

VI. Illustrative example

In this illustrative example the considered plant with two inputs and two outputs is described by following parameters

$$\boldsymbol{F} = \begin{bmatrix}0.9993 & 0.0987 & 0.0042\\ -0.0212 & 0.9612 & 0.0775\\ -0.3875 & -0.7187 & 0.5737\end{bmatrix}$$

$$\boldsymbol{G} = \begin{bmatrix}0.001 & 0.001\\ 0.0206 & 0.0197\\ 0.0077 & -0.0078\end{bmatrix}$$

$$\boldsymbol{C} = \begin{bmatrix}1 & 2 & 1\\ 0.15 & 1 & 1\end{bmatrix}$$

Constraint tying together chosen output variables is

$$\boldsymbol{D} = \begin{bmatrix}-3 & 2\end{bmatrix}$$

Using (19) and (20) matrices $\boldsymbol{M}$ and $\boldsymbol{N}$ were obtained

$$\boldsymbol{M} = \begin{bmatrix}16.3587 & -10.3787\\ 12.9918 & -8.2425\end{bmatrix}$$

$$\boldsymbol{N} = \begin{bmatrix}1 & 0 & -0.6132 & -0.4870\\ 0 & 1 & -0.4870 & -0.3868\end{bmatrix}$$

A. LMI approach

Following recomputed parameters were used in (23)

$$\varepsilon = 1.2$$

$$\boldsymbol{F}_1 = \begin{bmatrix}0.93018 & 0.05563 & 0.030254\\ -1.3402 & 0.13935 & 0.57466\\ 0.88766 & 0.075829 & 0.093071\end{bmatrix}$$

$$\boldsymbol{G}_1 = \begin{bmatrix}0.001 & 0.001 & 0.00020706 & -0.00050848\\ 0.0206 & 0.0197 & 0.003951 & -0.0097027\\ 0.0077 & -0.0078 & -0.0038196 & 0.0093801\end{bmatrix}$$

to obtain feedback matrix

$$\boldsymbol{K} = \begin{bmatrix}-47.5192 & 65.3576\\ 116.5412 & -160.5961\end{bmatrix}$$

and according to (33) it is possible to obtain signal gain matrix

$$\boldsymbol{W}_w = \begin{bmatrix}42.1803 & -11.5477\\ -11.7431 & -13.5503\end{bmatrix}$$

Designed controller is able to asymptotically stabilize given system as is shown in figure (1). Desired output variables are nonzero vector signal that is changed after output variables achieve desired values. The prescribed ratio of output variables is preserved and achieved before values of output variables reach desired value. Values of the desired output vector signal are in the defined ratio.

B. Optimal control approach

It is possible to change values of the following matrices $\boldsymbol{Q}$, $\boldsymbol{R}$ and $\boldsymbol{S}$ to obtain desired result

$$\boldsymbol{Q} = \begin{bmatrix}0 & 0\\ 0 & 0\end{bmatrix}$$

$$\boldsymbol{R} = \begin{bmatrix}0.0015 & 0\\ 0 & 0.0015\end{bmatrix}$$

$$\boldsymbol{S} = \begin{bmatrix}0.0015 & 0\\ 0 & 0.0015\end{bmatrix}$$

Following recomputed parameters were used in Matlab function `dare`.

$$\boldsymbol{F}^\circ = \begin{bmatrix}-0.68757 & 0.97503\\ -1.0314 & 1.4625\end{bmatrix}$$

$$\boldsymbol{G}^\circ = \begin{bmatrix}0.0499 & 0.0326 & -0.046476 & -0.036911\\ 0.02845 & 0.01205 & -0.023315 & -0.018516\end{bmatrix}$$

$$\boldsymbol{Q}^\circ = \begin{bmatrix}419.3882 & -279.9742\\ -279.9742 & 186.4441\end{bmatrix}$$

$$\boldsymbol{R}^{\circ} = 10^{-2} \begin{bmatrix} 1 & 0 & -0.61322 & -0.48701 \\ 0 & 1 & -0.48701 & -0.38678 \\ -0.61322 & -0.48701 & 0.61322 & 0.48701 \\ -0.48701 & -0.38678 & 0.48701 & 0.38678 \end{bmatrix}$$

$$\boldsymbol{S}^{\circ} = \begin{bmatrix} -16.069 & -12.7021 & 16.04 & 12.7386 \\ 10.6683 & 8.5322 & -10.6974 & -8.4956 \end{bmatrix}$$

Following feedback matrix is obtained.

$$\boldsymbol{K} = \begin{bmatrix} 5.6451 & -3.2698 \\ 4.5084 & -2.5591 \end{bmatrix}$$

In case of forced mode following signal gain matrix was used.

$$\boldsymbol{W}_w = \begin{bmatrix} 30.5887 & -3.2952 \\ -19.5852 & -8.4218 \end{bmatrix}$$

As is shown in figure (2) designed controller is able to asymptotically stabilize given system. Desired constraint defined as ratio of output variables is achieved before values of output variables reach desired value.

Desired output variables are nonzero vector signal and the signal is changed after output variables achieve desired values. The prescribed ratio of output variables is preserved. Values of the desired output vector signal are in the defined ratio.

VII. Conclusion

The considered problem of design of stabilizing control for linear discrete-time systems with equality constraint of output variables is approached by two different algorithms. First algorithm is LMI based algorithm, second one uses optimal control approach. The design conditions are prescribed as equality constraint tying together chosen output variables.

Well known property of SOF is inability in particular cases to find feedback matrix able to stabilize given system. Similarly, it is possible to determine feedback matrix using proposed approach unable to asymptotically stabilize given linear system and satisfy defined constraint, while different constraint can lead to acceptable results.

Both proposed algorithms are appliable and able to fulfil requirements of constraint defined as ratio of output variables.

ACKNOWLEDGEMENTS

The work presented in this paper was supported by VEGA, Grant Agency of Ministry of Education and Academy of Science of Slovak Republic under Grant No. 1/0256/11, as well as by Research & Development Operational Programme Grant No. 26220120030 realized in Development of Center of Information and Communication Technologies for Knowledge Systems. These supports are very gratefully acknowledged.

References

[1] D. Krokavec and A. Filasová, "Equality constraints in sensor faults reconfigurable control design", Proceedings of the 8th European Workshop on Advanced Control and Diagnosis ACD 2010, Ferrara, Italy, 2010, pp. 184-189.

[2] D. Krokavec and A. Filasová, "Constrained LQ control systems", In: Proceedings of the 16th International Conference Process Control 2007, June 11-14, 2007, Štrbské Pleso, Slovak Republic.

[3] V. Veselý: Static output feedback controller design. Kybernetika 57 (2001), 205-221.

[4] D. Rosinová, V. Veselý, V. Kučera: A necessary and sufficient condition for static output feedback stabilizability of linear discrete-time systems. Kybernetika 39 (2003), 447-459.

[5] B. Boyd, L. El Ghaoui, E. Peron, and V. Balakrishnan, Linear Matrix Inequalities in System and Control Theory. Philadelphia, SIAM Society for Industrial and Applied Mathematics, 1994.

[6] L. Vandenberghe, V. Balakrishnan: "Algorithms and software tools for LMI problems in control: An overview" In: Proceedings of the IEEE International Symposium on Computer-Aided Control system Design, Dearbora, Michigan, 1996

[7] D. Peaucelle, D. Henrion, Y. Labit, and K. Taitz, User's Guide for SeDuMi Interface 1.04, Toulouse: LAAS-CNRS, 2002

[8] D. Krokavec and A. Filasová, Dynamic Systems Diagnosis, Košice, Elfa, 2007. (in Slovak)

[9] D. Krokavec, A. Filasová, Diskrétne systémy, Košice : Elfa, 2006, ISBN 80-8086-028-9. (in Slovak)

[10] D. Krokavec, A. Filasová, V. Hladký, Residual generator design for a class of nonlinear systems described by Takagi-Sugeno models, In. Aspects of Computational Intelligence: Theory and Applications.
D. Krokavec, A. Filasová, V. Hladký, "Stabilizing fuzzy control for a class of nonlinear systems" In: Proceedings of the 10th IEEE Jubilee International Symposium on Applied Machine Intelligence and Informatics SAMI 2012, January 26-28, 2012, Herľany, Slovak Republic, [CD ROM] Óbuda University, Budapest, Hungary, s. 53-58. ISBN: 978-1-4577-0195-5, IEEE Catalog Number: CFP1208E-CDR (SCOPUS)

[11] P. Liščinský and V. Serbák, "Control of discrete-time linear systems constrained in output by equality constraints", In: Proceedings of the 14th International Carpathian Control Conference 2013, May 26-29,2013, Rytro, Poland. (Submitted)

[12] A. Filasová, V. Serbák and P. Liščinský, "Optimal control of discrete-time linear systems with output equality constraints", In: Proceedings of the 16th International Conference Process Control 2013, May 26-29, 2013, Strbské Pleso, Slovak Republic. (Submitted)

Performance analysis of the production line

[1]*Matej ČOPÍK (3rd year)*
Supervisor: [2]Ján JADLOVSKÝ

[1,2]Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

[1]matej.copik@tuke.sk, [2]jan.jadlovsky@tuke.sk

***Abstract* — This article deals with the performance analysis of the production line, which is modeled using Petri nets. The aim of this paper is to present a methodology that allows the performance analysis of real production lines or even the lines in their design phase. A current trend in the design of production lines is their modeling and analysis. This approach ensures the required properties and the possibility of preparing a production planning. Possibility to plan their production in the required quantity and quality on a fixed time or on periodic intervals is a goal of many manufacturing plants today.**

***Keywords* — Petri nets, manufacturing line, performance, analysis**

I. Introduction

Production systems in present are different from the older mainly in their complexity and the technology used. Many production plants are using fully or semi-automated production systems with high requirements on reliability, quality, productivity and efficiency. One of the reasons for the fully automated production is to reduce the production time and thereby increasing the production. Increasing the production is connected with the increasing of requirements on storage space if the production logistics is missing. In such cases the JIT (just in time) systems come to the fore, which are designed to produce just right quantity at the time required and there is no need for storing. Analysis of the individual parts and mainly of the connections among them is the important part of the design when using JIT systems. An analysis of the project in its design phase often provides suggestions for improvement that are easily incorporated into the project. Analysis of the real systems can sometimes bring better results, but the improvements are difficult. In the analysis with respect to the JIT system is necessary to consider the real production capacity, which often differs from the ideal one of 20 to 30%. In modeling and later in the analysis of production systems in the design phase is often overlooked the fact that the results are often overstated and in practice unattainable. For modeling and analysis of the production line the Petri nets (PN) were chosen, which are suitable for discrete system modeling. Modeling allows us to know the characteristics of the production system before implementing or before modification. The simulation results are helpful in verifying the required performance of the production system.

First part of the article is devoted to the description of the production line model, which is modeled by timed PN.

The second part is devoted to the analysis of this model. The results of the model analysis are compared with the experimental results obtained on the real production line.

II. Production Line Description

A. Description of the production line model

Production line can be also considered as the model of fully automated production line (Fig. 1), which was named Flexible Manufacturing System (FMS) and is used for educational and research purposes.

Fig. 1. Flexible Manufacturing System

The model is located at the Department of Cybernetics and Artificial Intelligence (DCAI) of the Technical University (TU) in Košice. The model is located at the laboratory V147, on Vysokoškolska Street number 4. FMS is used in the learning process and in the process of bachelor, master and

doctoral theses creation. This model is divided into 5 posts that represent different parts of the production process. The manufacturing process starts with the templates selection from the stock, which is formed by the input and output storage. The selected template is placed on a conveyor belt by the manipulator, and moves to the end of the belt. It is then transferred to second conveyor belt, after which it moves to the fifth post. If the template is full it is emptying on this post. Empty template is moved to the post 1, where it is stored in with the colored cubes. This work presents an accurate assembly of parts or materials for a specified location on the product. After storing, the template moves on the end of the second conveyor, where the video camera control is made. It is in charge of verifying the correctness of the pattern stored. Finally, the template is moved back to the first conveyor, moved to the post 3 and stored into the output storage. This entire production cycle is performed automatically and there can be simultaneously up to four templates in the production process. Each post is managed independently and its activity begins with the template arrival. This allows parallel handling up to four templates. Moreover, the sharing of resources is present on the post 3, where can the requirements for entry and exit of the template occur at one moment. A more detailed description can be found in [1] and [3].

B. Description of the FMS model using PN

FMS model described in the previous section is modeled by a generalized PN, which is shown in Fig.2. This model describes the activities of FMS.

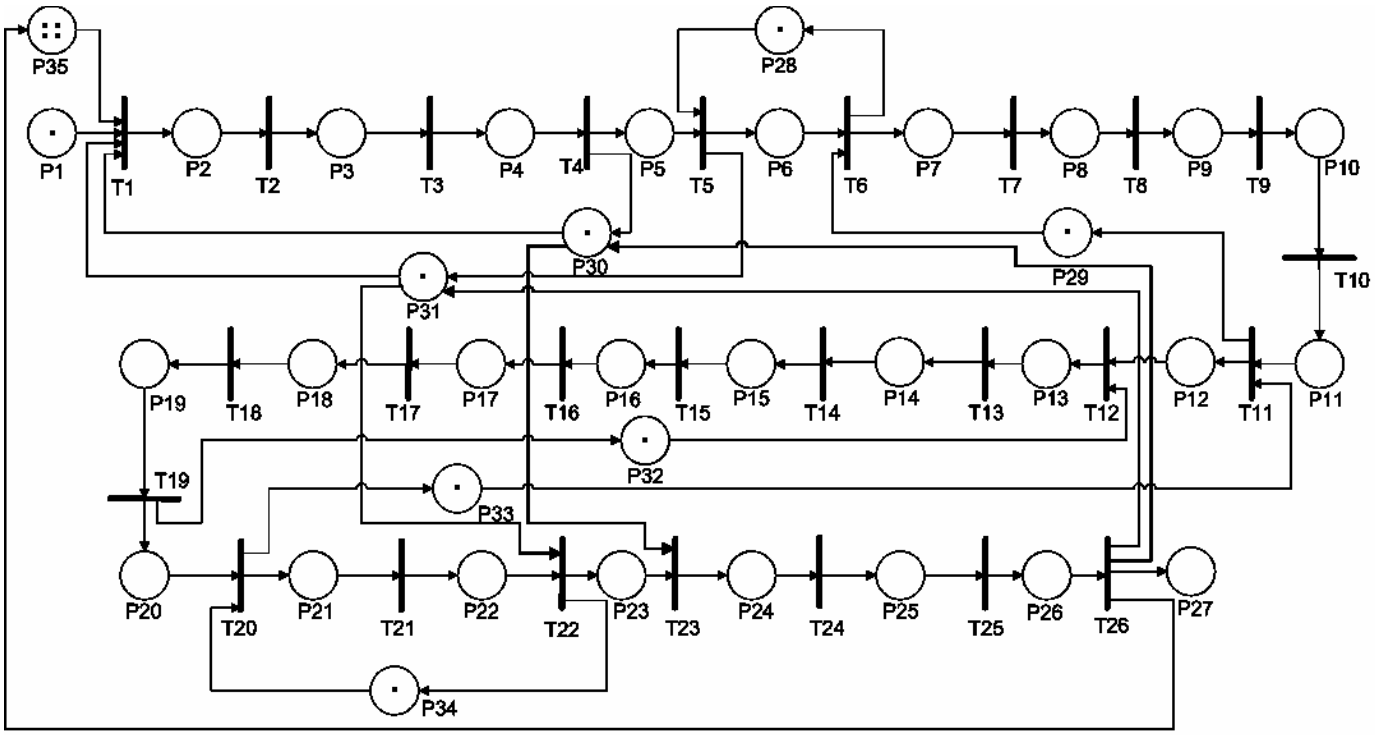


Fig. 2. PN model of the FMS.

Places P1 to P27 and transitions T1 to T26, is a major production of the production line where each operation is performed as described in Chapter II. Places P28, P29, P31, P33 and P34, determine whether the post is free, or the operation is held. Places P30 and P34 determine readiness of the manipulators. Place P35 determines the number of templates that can be produced at one time. The number of templates is limited to a maximum of four templates in the production. The initial marking of the model is derived from the state of the production line. All positions are free and manipulators are ready for the production. On the input stock P1 is one token representing a requirement to one template production. In the place P35 are located four tokens, providing an opportunity to produce four pieces at a time.

For the performance analysis of this model it is necessary to extend the model for times of the operations. These times can be obtained in the design phase from the specifications of the equipment intended to be used for the production line implementation. In this case, the individual times were obtained experimentally from the real model in ideal conditions. In order to obtain the time data there were performed 100 measurements each time and the average value was considered for the reference one. Expending the generalized PN with the times of individual transitions we obtain the T - timed PN. The model of timed PN will be used to analyze the performance of the manufacturing line, which is compared to a real production line under standard conditions. Ideal conditions compared to the standard ones always provide full cube stores, constant pressure, and failure-free state.

The proposed model of the production line must satisfy certain properties of PN such as reachability, liveliness, conflictless, and can not include own cycles.

PN in Fig. 2 can be described using matrices to describe T - timed PN.

$$N = (\boldsymbol{P}, \boldsymbol{T}, \boldsymbol{pre}, \boldsymbol{post}, \boldsymbol{M_0}, \boldsymbol{Tempo}) \quad (1)$$

where $\boldsymbol{P} = \{p1, p2, ..., pm\}$ is a finite set of places,
$\boldsymbol{T} = \{t1, t2, \ldots, tn\}$ is a finite set of transitions,
pre - is direct incidence function, $\boldsymbol{pre}: \boldsymbol{P} \times \boldsymbol{T} \rightarrow N$,

post – is backward incidence function, ***post***: $\boldsymbol{P} \times \boldsymbol{T} \rightarrow N$,
$\boldsymbol{M_0}$ – is initial marking $\boldsymbol{M_0}: \boldsymbol{P} \rightarrow N$,
Tempo $(t_i) = d_i$ – specifies the time interval of the implementation of transitions, where $d_i \geq 0$.

For the analysis it is necessary to know the matrixes ***post*** and ***pre*** from which it is possible to calculate the incidence matrix ***C*** [4] according to equation (2).

$$\boldsymbol{C} = \boldsymbol{post} - \boldsymbol{pre} \tag{2}$$

Incidence matrix for the PN in Fig. 2 is not presented according to its dimensions of 54 x 26. Using the matrix C, M_0 and M_X, which represents the desired state of the system, it is possible to work with the PN model not only graphically, but also analytically.

III. Performance analysis of FMS

Performance analysis is created on the FMS model described in the previous chapter. Times of individual transitions have been obtained experimentally. These times can be entered into the rate matrix, which indicates the time intervals of the transitions. In the production systems are in terms of their analysis of the most interesting hour performance, it is how many pieces can produce per hour. In the case of fully automated production lines, it is interesting to see how many pieces produce for one shift, or 8 hours. For comparison, we create the analysis and the production time of one and hundred of products.

As the entry requirements for the analysis we consider:

1. number of pieces produced per 1 hour
2. number of pieces produced per 8 hour
3. production time of one piece
4. production time of 100 pieces

Analysis of the production line can be created in programs for modeling and analysis of PN such as CPNtool. But this program does not offer the time unit conversion (in this case milliseconds), into the form hh:mm:ss.fff, so the analysis is difficult. It also does not allow further processing of the results. The analysis was therefore created in the PNA tool described in [2], whose output is the time value in the form hh:mm:ss.fff and there is possibility of further work with the results.

After production line analysis, we can say that in the case of 100% performance it is able to produce:

1. 80 pieces in 1 hour
2. 643 pieces in 8 hours
3. 1 product for 45 second and 591 thousandths
4. 100 products for 1 hour 14 minutes 36 seconds and 534 thousandths

These values are valid for 100% performance of the production line.

The real production systems have certain requirements to the operation of the device. Operation of the device consists of activities during which the manufacturing process is not on. For example: insufficient supply of media, debugging, adding of missing material to storage, maintenance of the manufacturing process, waiting for the primary component, etc. Ideal hourly performance of each production system actually decreases for the following reasons. According to [5], it is generally assumed ideal hourly decrease from 20% to 30% and even more in some of the production processes. Therefore, it is necessary to calculate the real performance per hour, which represents 80% or 70% of the ideal hourly performance of the manufacturing process. Table Tab. 1 summarizes the results of the simulation at 100%, 80% and 70% performance of production line (time is shown in the form h:mm:ss.fff).

Performance	100%	80%	70%
Number of pieces per 1 hour	80	64	56
Number of pieces per 8 hour	643	514	450
Production time of 1 piece	0:0:45.591	0:0:54.709	0:0:59.268
Production time of 100 pieces	1:14:36.534	1:29:31.841	1:36:59.494

Tab. 1.: Summary of simulation results

To verify of the simulation results, there were carried out experimental measurements on the real system, which was described in chapter II, and the simulation model was also based on it. During the measuring on the real system, there was an effort to simulate real conditions. In real measurements a few errors were simulated on the line. Like occasional waiting for the missing cube in the stack, occasional loss of pressure or simulation of line fault. All of these situations happen in real productions; therefore it was necessary to enforce the errors into the system during the control measurements on the real system. The results of experimental measurements are reported in table Tab. 2. In experiments 1 and 4 there were performed 5 measurements, and the result is the average value of these measurements. Experiment 2 was performed only once, because it is time-consuming. Experiment 3 was performed 50 times and the result is average value of these measurements.

	Average time / pc	Performance [%]
Number of pieces per 1 hour	67	83,75
Number of pieces per 8 hour	554	86,16
Time of 1 piece	0:0:56.389	80,08
Time of 100 pieces	1:28:11.965	84,59

Tab. 2.: The results from the real system

By comparing the experimental results obtained from the simulation results it can be concluded that the performance of the real production line is in ranges from 80.08 to 86.16%. These results correspond to the reality, when the error of lower air pressure was present during the measurements of one piece production, but for 8 hour measurement, this error did not occur all the time. Measurement has also showed that some of the lines actuators can be set better. The results of the additional measurements and analysis will be the subject of

other articles, since it requires more complex optimization of production line actuators.

IV. Conclusion

Progressive methods used to design and implement of production lines are more and more used in practice. Modeling and analysis is of great importance for the demanding performance requirements of the system. This work provides the use of PN as one of the ways to model and analysis of production lines. The results presented in this article suggest a potential for the analysis of the production line possible also in the design process.

Comparing simulation and experimental measurements results shows that the performance of production line in average is 82.65%. This performance compared to the performance considered in the literature (70% to 80%) can be considered as sufficient and it can be further improved by optimizing the actuators settings used in the production line. Without the use of the model and its analysis, we have might never come to the conclusion that the production line is able to work faster. Therefore it is possible to say that the analysis of the production line model helped to improve the performance of the real production line. In general we can say that the analysis of the model answers to the central question of whether the production system meets the required performance or not.

Acknowledgment

This work has been supported by the Scientific Grant Agency of Slovak Republic under project Vega No.1/0286/11 Dynamic Hybrid Architectures of the Multiagent Network Control Systems.

References

[1] ČOPÍK, M. – LACIŇÁK, S. – JADLOVSKÝ, J. 2010. Návrh a realizácia riadenia pružnej výrobnej linky. In Electrical Engineering and Informatics: Proceeding of the Faculty of Electrical Engineering and Informatics of the Technical University of Košice [CD-ROM]. Košice : Technická univerzita, 2010 s. 561-564. ISBN 978-80-553-0460-1.

[2] ČOPÍK, M – ILKOVIČ, J. Analysis of manufacturing systems using Petri nets. In: SCYR 2012 : proceedings from conference : 12th Scientific Conference of Young Researchers : May 15th, 2012, Herľany, Slovakia. - Košice : TU, 2012 S. 72-75. - ISBN 978-80-553-0943-9

[3] Flexible Manufacturing System, dostupné na internete: http://kyb.fei.tuke.sk/laben/modely/fmp.php

[4] KORDIC, V. 2008. Petri net theory and applications. Vienna: I-Tech Education and Publishing, 2008. 544 s. ISBN 978-3-902613-12-7

[5] MADARÁSZ, L. 2006. Integračné aspekty tvorby a prevádzky systémov CIM : 2.vydanie. Košice : Elfa, s.r.o., 2006. 380 s. ISBN 80-86-043-2

Predictive Control using the Nonlinear Predictor

[1]*Štefan JAJČIŠIN (3rd year)*
Supervisor: [2]*Anna JADLOVSKÁ*

Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

[1]stefan.jajcisin@tuke.sk, [2]anna.jadlovska@tuke.sk

***Abstract*—The main goal of the paper is to introduce the modification of basic predictive control principle with using the nonlinear predictor of controlled system's behaviour. The basic principle and design of predictive control algorithms based on the linear model is mentioned in the paper, too. Testing of introduced control algorithms is carried out by control of the laboratory hydraulic system.**

***Keywords*—hydraulic system, nonlinear dynamical system, nonlinear predictor, predictive control.**

I. Introduction

The paper deals with the predictive control with using a linear and a nonlinear predictor of controlled system's behaviour. Firstly, the basic principle and design of predictive control algorithms based on the linear model are introduced. Next the modification of control algorithms design using the nonlinear predictor is mentioned. As controlled system for algorithms testing the laboratory model of hydraulic system is used, whereby its predictive control based on the linear model was published in [8]. In this article only results obtained by predictive control with nonlinear predictor are presented.

II. Principle of Predictive Control

The typical feature of the model-based predictive control (MPC) is using the behaviour prediction of controlled physical system in control action computation at each sample. Predicted values of particular quantities are computed on the basis of the controlled system's model.

Regarding to used model of controlled system predictive control algorithms can be divided into two categories:

A. algorithms based on the linear approximation of nonlinear physical system – as discrete transfer function

$$F_s(z^{-1}) = \frac{B_z(z^{-1})}{A_z(z^{-1})} = \frac{\sum_{i=0}^{m} b_i z^{-i}}{\sum_{i=0}^{n} a_i z^{-i}}, \quad (1)$$

where $B_z(z^{-1})$ is numerator's polynom (order m, coefficients b_i) and $A_z(z^{-1})$ denominator's polynom (order n, coefficients a_i), or in discrete state-space model form

$$\begin{aligned} \boldsymbol{x}(k+1) &= \boldsymbol{A}_d \boldsymbol{x}(k) + \boldsymbol{B}_d \boldsymbol{u}(k) \\ \boldsymbol{y}(k) &= \boldsymbol{C}\boldsymbol{x}(k) + \boldsymbol{D}\boldsymbol{u}(k) \end{aligned}, \quad (2)$$

where $\boldsymbol{A}_d$, $\boldsymbol{B}_d$, $\boldsymbol{C}$ and $\boldsymbol{D}$ are matrices, $\boldsymbol{x}(k)$ is state vector, $\boldsymbol{u}(k)$ is input vector and $\boldsymbol{y}(k)$ is vector of system's output.

B. algorithms, which use the nonlinear model like

$$\begin{aligned} \dot{\boldsymbol{x}}(t) &= \boldsymbol{f}\left(\boldsymbol{x}(t), \boldsymbol{u}(t), t\right) \\ \boldsymbol{y}(t) &= \boldsymbol{g}\left(\boldsymbol{x}(t), \boldsymbol{u}(t), t\right) \end{aligned}, \quad (3)$$

where $\boldsymbol{x}(t)$ is state vector, $\boldsymbol{u}(t)$ is input vector, $\boldsymbol{y}(t)$ is vector of system's output, $\boldsymbol{f}$ and $\boldsymbol{g}$ are vector nonlinear functions.

In predictive control algorithms, an optimization task is executed for computing the value of control action. Its main principle consists in minimization of criteria function

$$J_{MPC} = \sum_{i=N_1}^{N_p} \boldsymbol{Q}_e \left[\hat{\boldsymbol{y}}(k+i) - \boldsymbol{w}(k+i)\right]^2 + \sum_{i=1}^{N_u} \boldsymbol{R}_u \left[\boldsymbol{u}(k+i-1)\right]^2, \quad (4)$$

where $\hat{\boldsymbol{y}}(k)$ is vector od predicted output, $\boldsymbol{w}(k)$ is vector of desired value, $\boldsymbol{u}(k)$ is vector of control action values [1]. Values N_1 and N_p represent the prediction horizon and N_u constitutes the control horizon, on which the optimal sequence of control action $\boldsymbol{u}(k)$ is computed, whereby $N_u \leq N_p$ [2].

The *Receding Horizon Strategy* is typical for predictive control algorithms [1]. It means the optimal sequence of control action $\boldsymbol{u}_{opt} = \left[\boldsymbol{u}_{opt}(k) \ \cdots \ \boldsymbol{u}_{opt}(k+N_u)\right]$ is computed on control horizon N_u at each sample k, however only the first element $\boldsymbol{u}_{opt}(k)$ is used as system's input $\boldsymbol{u}(k)$.

The most used approach to predictive control algorithms design can be divided to three steps:

1) the predictor derivation on the basis of the controlled system's linear model,
2) the expression of gradient $\boldsymbol{g}^T$ and Hessian matrix $\boldsymbol{H}$,
3) the optimal sequence of control action computing by criteria function minimization.

We are focused on the MPC algorithms design's particular steps in next paper's part.

III. Design of MPC Algorithms with Linear Predictor

The input of the first step of MPC algorithms design is the linear model of controlled system. The result of this step is the predictor in the matrix form

$$\hat{\boldsymbol{y}} = \boldsymbol{y}_f + \boldsymbol{G}\Delta\boldsymbol{u}, \quad (5)$$

where $\hat{\boldsymbol{y}}$ is vector of predicted output values, $\boldsymbol{y}_f$ is vector of system's free response and $\boldsymbol{G}\Delta\boldsymbol{u}$ constitutes the system's forced response [1].

The concrete expression of vector $\boldsymbol{y}_f$ and matrix $\boldsymbol{G}$ depends on used form of the physical system's linear model. We are using the state-space model of dynamical systems in this paper, where the control action rate can be written explicitly:

$$\begin{aligned} \boldsymbol{x}(k+1) &= \boldsymbol{A}_d\boldsymbol{x}(k) + \boldsymbol{B}_d\boldsymbol{u}(k-1) + \boldsymbol{B}_d\Delta\boldsymbol{u}(k) \\ \boldsymbol{y}(k) &= \boldsymbol{C}\boldsymbol{x}(k) + \boldsymbol{D}\boldsymbol{u}(k) \\ \boldsymbol{u}(k) &= \boldsymbol{u}(k-1) + \Delta\boldsymbol{u}(k) \end{aligned} \quad . \tag{6}$$

According to [3] it is possible to derive the predictor in form

$$\hat{\boldsymbol{y}} = \underbrace{\boldsymbol{V}\boldsymbol{x}(k) + \boldsymbol{G}_1\boldsymbol{u}(k-1)}_{\boldsymbol{y}_f} + \boldsymbol{G}\Delta\boldsymbol{u}\,, \tag{7}$$

by iterations of state-space model equations (3). In formula (7) the free response $\boldsymbol{y}_f$ is computed on the basis of state quantities current values $\boldsymbol{x}(k)$ and input's previous values $\boldsymbol{u}(k\text{-}1)$, where provided that $\boldsymbol{D}$ is matrix of zeros

$$\boldsymbol{V} = \begin{pmatrix} \boldsymbol{C}\boldsymbol{A}_d \\ \vdots \\ \boldsymbol{C}\boldsymbol{A}_d^{Np} \end{pmatrix},\ \boldsymbol{G}_1 = \begin{pmatrix} \boldsymbol{C}\boldsymbol{B}_d \\ \boldsymbol{C}\left(\boldsymbol{A}_d + \boldsymbol{I}\right)\boldsymbol{B}_d \\ \vdots \\ \boldsymbol{C}\left(\boldsymbol{A}_d^{Np-1} + \cdots + \boldsymbol{A}_d + \boldsymbol{I}\right)\boldsymbol{B}_d \end{pmatrix}, \tag{8}$$

$$\boldsymbol{G} = \begin{pmatrix} \boldsymbol{C}\boldsymbol{B}_d & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{C}\left(\boldsymbol{A}_d^{Np} + \boldsymbol{I}\right)\boldsymbol{B}_d & \boldsymbol{C}\boldsymbol{B}_d & \ddots & \vdots \\ \vdots & \ddots & \ddots & \boldsymbol{0} \\ \boldsymbol{C}\left(\boldsymbol{A}_d^{Np-1} + \cdots + \boldsymbol{A}_d + \boldsymbol{I}\right)\boldsymbol{B}_d & \cdots & \boldsymbol{C}\left(\boldsymbol{A}_d + \boldsymbol{I}\right)\boldsymbol{B}_d & \boldsymbol{C}\boldsymbol{B}_d \end{pmatrix}$$

It is needed to reduce the matrix $\boldsymbol{G}$ regarding to the length of control horizon N_u

$$\boldsymbol{G} \leftarrow \boldsymbol{G}\begin{pmatrix} \boldsymbol{I} \\ \boldsymbol{0} \end{pmatrix}, \tag{9}$$

where $\boldsymbol{I}$ is unit matrix with dimension $N_u.n_u$, whereby n_u is number of system's inputs[1].

The second step of MPC algorithms design is focused on the work with criteria function's (4) in matrix form – in our case we consider the weighting of control action rate $\Delta\boldsymbol{u}$

$$J_{MPC} = \left(\hat{\boldsymbol{y}} - \boldsymbol{w}\right)^T \boldsymbol{Q}\left(\hat{\boldsymbol{y}} - \boldsymbol{w}\right) + \Delta\boldsymbol{u}^T\boldsymbol{R}\Delta\boldsymbol{u}\,. \tag{10}$$

It should be expressed in suitable quadratic form for computing the optimal sequence of control action rate $\Delta\boldsymbol{u}$. The particular vectors $\hat{\boldsymbol{y}}$, $\boldsymbol{w}$, $\Delta\boldsymbol{u}$, and weighing matrices $\boldsymbol{Q}$, $\boldsymbol{R}$ have their dimensions in accordance with the length of horizons N_p, N_u and numbers of controlled system's outputs and inputs.

After the predictor (7) substitution to the criteria function (10) we can obtain the equation

$$\begin{aligned} J_{MPC} &= \Delta\boldsymbol{u}^T\left(\boldsymbol{G}^T\boldsymbol{Q}\boldsymbol{G} + \boldsymbol{R}\right)\Delta\boldsymbol{u} + \\ &+ \left[\left(\boldsymbol{y}_f - \boldsymbol{w}\right)^T\boldsymbol{Q}\boldsymbol{G}\right]\Delta\boldsymbol{u} + \Delta\boldsymbol{u}^T\left[\boldsymbol{G}^T\boldsymbol{Q}\left(\boldsymbol{y}_f - \boldsymbol{w}\right)\right] + c, \end{aligned} \tag{11}$$

which can be expressed as quadratic form

$$J = \Delta\boldsymbol{u}^T\boldsymbol{H}\Delta\boldsymbol{u} + 2\boldsymbol{g}^T\Delta\boldsymbol{u} + c\,, \tag{12}$$

where the matrix $\boldsymbol{H}$ and the vector $\boldsymbol{g}^T$ are

$$\begin{aligned} \boldsymbol{H} &= \boldsymbol{G}^T\boldsymbol{Q}\boldsymbol{G} + \boldsymbol{R}, \\ \boldsymbol{g}^T &= \left(\boldsymbol{y}_f - \boldsymbol{w}\right)^T\boldsymbol{Q}\boldsymbol{G}. \end{aligned} \tag{13}$$

The minimization of quadratic form (12) is executed in the third step of MPC algorithms design. In opposite to classical dynamical systems control approaches the advantage of MPC algorithms is possibility to respect system's constraints in optimal sequence of control action computing.

We are using the function *quadprog*, which is part of the *Optimization Toolbox* in Matlab for quadratic form (12) minimization. The algorithm for vector of optimal values $\Delta\boldsymbol{u}$ computing by formula

$$\min_{\boldsymbol{u}} \frac{1}{2}\Delta\boldsymbol{u}^T\boldsymbol{H}\Delta\boldsymbol{u} + \boldsymbol{g}^T\Delta\boldsymbol{u}\,,$$

with respect to $\boldsymbol{A}_{con}\Delta\boldsymbol{u} \le \boldsymbol{b}_{con}$, (14)

is implemented in *quadprog* function.

The matrix $\boldsymbol{A}_{con}$ and the vector $\boldsymbol{b}_{con}$ should be defined according to system's constraints [2].

While the predictor derivation and Hessian matrix $\boldsymbol{H}$ expression can be executed in advance, the values of gradient $\boldsymbol{g}^T$ and criteria function $\boldsymbol{J}_{MPC}$ minimization are carried out during control loop. This approach to predictive control is the most used all over the world. We were considered it in control of the Hydraulic laboratory model in [8].

IV. Design of MPC algoriths with nonlinear predictor

It is obvious from text before, that the control action value is computed on the basis of the linear model at each sample time in MPC algorithms. Thus, it is very important to approximate the controlled system's dynamics good enough by the linear model. In opposite case the control cannot be adequate, eventually it can destabilize the controlled system.

In this paper's part we present a modified approach to MPC algorithms, where the predictor with nonlinear character is used in computing the control action. As predictor we will use nonlinear differential equations (3) – next only NDE, describing the controlled system's dynamic.

We are engaged in two variants of using the nonlinear predictor in this paper. In the first case, we will use the basic principle of MPC, where we will modify the design procedure in such a case that the system's free response will be computed from the solution of NDE. Any other computations will be based on the predictor in linear form. The second variant is using the fully nonlinear predictor.

Using the nonlinear predictor in free response computing

Regarding to the nonlinear character of model, the computing of system's free response vector values $\boldsymbol{y}_f$ is possible only by numerical methods. For NDE equations (3) we will use solution by 4th ordered Runge-Kutta method [4]. It is necessary to execute the prediction cyclically for N_p samples, where the system's state from sample $k-1$ is used as initial values for computing the solution in current sample k.

In programming way, it is also necessary to modify the function for control action computing in accordance with Fig. 1. In the frame of control action computing, there are only few more numerical computations in control loop in comparison to classical approach with completely linear predictor. Therefore, we suppose only little, possibly neglectable increasing of MPC algorithm's computational time. The advantage of this approach is that the minimization of criteria function, what is the critical part of MPC algorithms regarding to computational time, stays unchanged. It means the optimization problem can be rewritten to quadratic form and computed by the numerical method of quadratic programming. As we have already mentioned it, the nonlinear predictor in system's free response computing constitutes only partial using of predictor's nonlinear character. The nonlinear character affects only values of gradient $\boldsymbol{g}^T$.

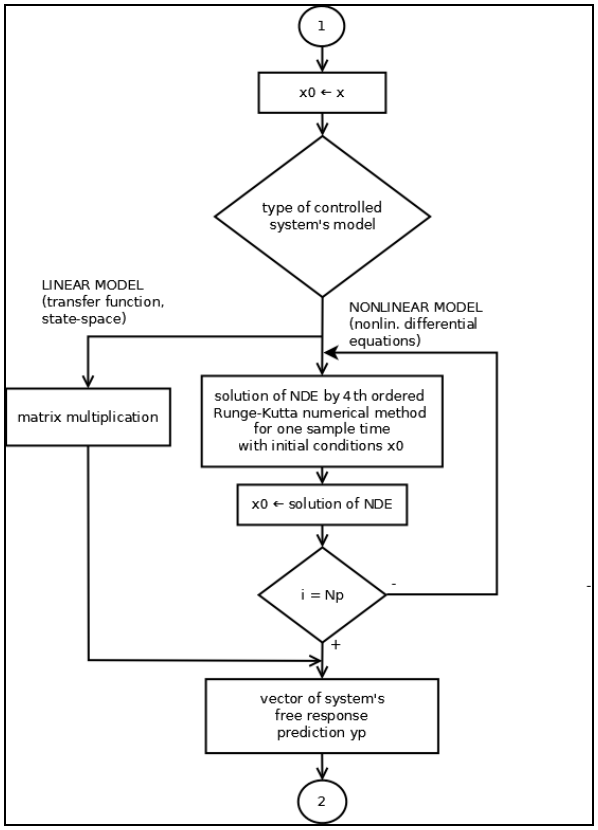


Fig. 1 Part of flow chart diagram for computing the controlled system's free response

In case of systems, which cannot be approximated by the linear model good enough, even this approach cannot be useful to ensure desired behaviour of controlled system. Therefore, in next paper's part we will focused on the nonlinear predictive control, which uses the nonlinear predictor in whole MPC algorithms design.

Nonlinear predictive control

The nonlinear predictive control keeps the basic principle of MPC, however the main idea of this approach is that the model and the predictor are defined by nonlinear functions $\boldsymbol{f}$ a $\boldsymbol{g}$:

$$\begin{aligned} \boldsymbol{x}(k+1) &= \boldsymbol{f}\left[\boldsymbol{u}(k), \boldsymbol{x}(k), \boldsymbol{v}(k), \boldsymbol{z}(k)\right] \\ \boldsymbol{y}(k) &= \boldsymbol{g}\left[\boldsymbol{x}(k)\right] + \boldsymbol{\xi}(k) \end{aligned}, \tag{15}$$

where $\boldsymbol{u}(k)$ is vector of inputs, $\boldsymbol{x}(k)$ is vector of state quantities, $\boldsymbol{v}(k)$ is vector of measurable disturbances, $\boldsymbol{z}(k)$ is vector of not measurable disturbances, $\boldsymbol{\xi}(k)$ is noise vector and $\boldsymbol{y}(k)$ is vector of system's outputs [5].

It is not possible to use the MPC algorithms design, which was introduced in previous paper's parts in this case, because it is not able to express the predictor in matrix form (7) from equations (15). Thus, it is not possible to rewrite the criteria function in quadratic form (12). For that reason it is necessary to do direct minimization of criteria function by numerical methods for computing the control action value $\boldsymbol{u}(k)$. In our case we used method of *Sequence Quadratic Programming* (SQP). According to [6] it can be proved that solution obtained by SQP is equivalent with Newton-Lagrange solution and it converge quadratically near the minimum point.

We used function *fmincon,* which is part of *Optimization Toolbox* in Matlab for obtaining the nonlinear optimization problem solution. By suitable choice of optimization algorithm (in our case it is SQP) it is possible to find a minimum of arbitrary function *Fun* with respect to constraints defined by $\boldsymbol{A}_{con}$, $\boldsymbol{b}_{con}$ near to the point $\boldsymbol{u}_0$. The simplest syntax of this function in Matlab is

$$\boldsymbol{u} = fmincon(Fun, \boldsymbol{u}_0, \boldsymbol{A}_{con}, \boldsymbol{b}_{con}). \tag{16}$$

In case of nonlinear predictive control the function *Fun* represents the criteria function (10), where computing of the predicted output values is carried out by numerical Runge-Kutta method again.

Principially, it is possible to express the nonlinear predictive control algorithm by Fig. 2.

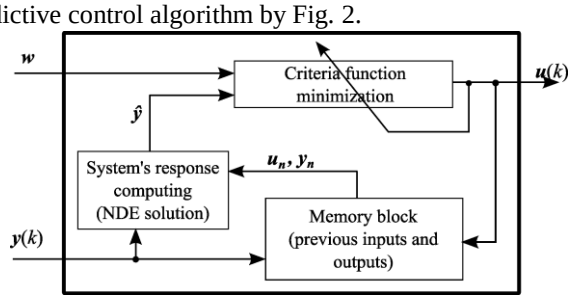


Fig. 2 Nonlinear predictive control algorithm

The minimization of criteria function in nonlinear form spends a lot of computational time, therefore the nonlinear predictive control should be used in control of dynamic systems with slower dynamics, mainly in control of thermal, chemical or hydraulic systems. Especially for that reason we applied the nonlinear predictive control to laboratory model of Hydraulic system, which is located in Laboratory of mechatronic systems at the Department of Cybernetics and Artificial Intelligence (http://kyb.fei.tuke.sk/laben/modely/hyd.php).

V. HYDRAULIC LABORATORY MODEL CONTROL WITH MPC ALGORITHMS WITH NONLINEAR PREDICTOR

In this part we are engaged in predictive control of Hydraulic laboratory model, where MPC algorithms with nonlinear predictor, introduced in part IV, are used. Regarding to the fact, that predictive control is primarily used for systems with slow dynamics, we used the hydraulic system. The predictive control of this system based on the linear predictor has already been carried out in [8].

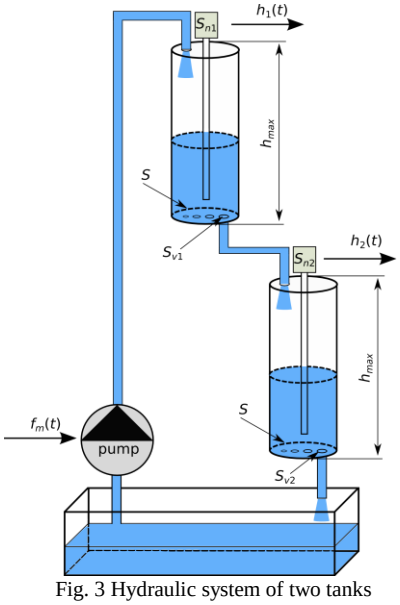


Fig. 3 Hydraulic system of two tanks

The hardware configuration and systemic view of hydraulic system was presented in [7]. Nonlinear differential equations, which represent the model's mathematically-physical description are:

$$\frac{dh_1(t)}{dt} = \frac{1}{S}\left(k_p U(t) - S_{v1}\sqrt{2gh_1(t)}\right)$$
$$\frac{dh_2(t)}{dt} = \frac{1}{S}\left(S_{v1}\sqrt{2gh_1(t)} - S_{v2}\sqrt{2gh_2(t)}\right) \quad (17)$$

where g is the acceleration of gravity and the value k_p expresses the relation between input voltage $U(t)$ and inflow $q_{in1}(t)$. A schematic illustration of hydraulic system is depicted in Fig. 3, whereby particular physical parameters are:

S - intersection of tanks,
S_{v1}, S_{v2} - intersection of outlets of both tanks,
h_{max} - height of tanks (maximal liquid level).

Physical quantities

$f_m(t)$ - pump's motor frequency,
$h_1(t)$, $h_2(t)$ - current levels of liquid in both tanks

constitute system's input and outputs.

Sensors, which scan the current liquid level in both tanks are marked as S_{n1} and S_{n2}.

We present results of Hydraulic system control with predictive control algorithms with nonlinear predictor on next figures. Results are presented as time responses of control action and liquid levels in both of tanks, whereby the goal of control was to ensure desired value of liquid level in the second tank $h_2(k)$. We used algorithms setting in accordance to Tab. 1, where T_s is the sampling time and $\boldsymbol{I}$ is a unit matrix.

T_s	N_p	N_u	$\boldsymbol{Q}$	$\boldsymbol{R}$	constraints
4s	10	2	$1000\boldsymbol{I}$	$0{,}01\boldsymbol{I}$	$u \in \langle 0;8 \rangle V$

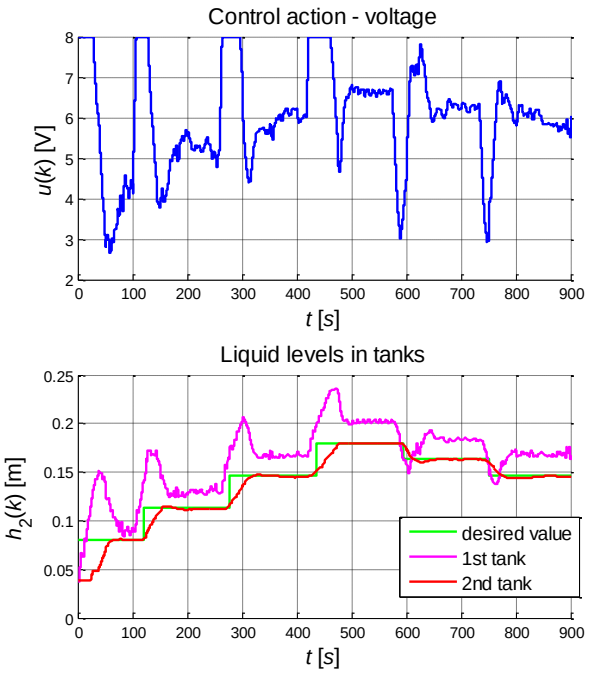


Fig. 4 Time responses of Hydraulic system laboratory model predictive control with nonlinear predictor of system's free response.

By comparison with results published in [8] we can allege that using the nonlinear predictor in dynamical systems predictive control brings better results than with the linear predictor. Especially in control action periodicity and overshooting the desired value by controlled quantity.

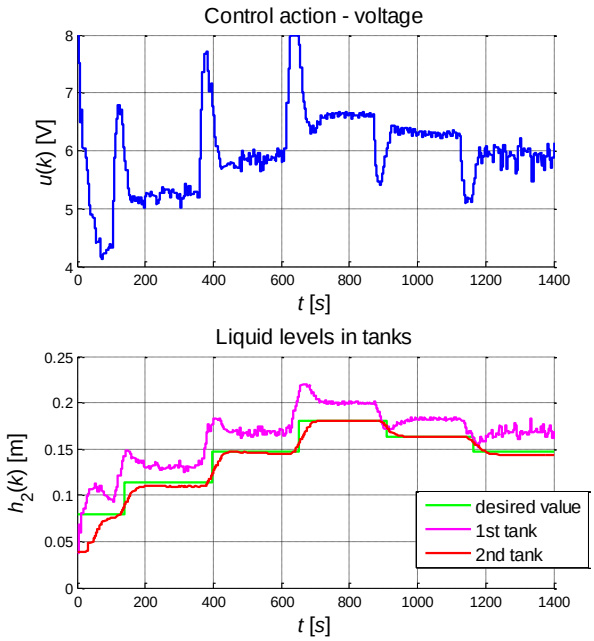


Fig. 5 Time responses of nonlinear predictive control of Hydraulic system laboratory model

VI. Conclusion

We mentioned the basic principle and design of predictive control algorithms with linear predictor. We also introduced two modifications of basic principle with using the nonlinear predictor, which we programmed and tested in laboratory model control. Based on presented results we can consider modified algorithms as suitable for dynamic systems control.

Acknowledgment

This work has been supported by the Scientific Grant Agency of Slovak Republic under project Vega No.1/0286/11 Dynamic Hybrid Architectures of the Multiagent Network Control Systems (70%). This work is also the result of the project KEGA 021TUKE-4/2012 (30%).

References

[1] J. Roubal, "Predictive controller", (in Czech), (Examples for exercises – Modern Control Theory). Available on the internet: <http://support.dce.felk.cvut.cz/pub/roubalj/teaching/MTR/seminars/MTR_cv8_mpc.pdf>.

[2] M. Fikar, "Predictive Control – An Introduction", Bratislava: Slovenská technická univerzita - FCHPT, 1999.

[3] K. Belda, J. Böhm, M. Valášek, "Model-based control for parallel robot kinematics", Proceedings of the 3rd International Congress on Mechatronics. MECH2K4, p. 1-15, 2004.

[4] L. F. Shampine, "Numerical Solution of Ordinary Differential Equations", Chapman & Hall, New York, 1994B.

[5] Kouvaritakis, M. Cannon, "Nonlinear Predictive Control, theory and practice," IET Control Engineering series 61, ISBN 978-0-85296-984-7

[6] R. Fletcher, "Practical methods of optimization, second edition", Wiley, 2000, ISBN 978-0471494638.

[7] Š. Jajčišin, "Verification of Control Algorithms with DDE Communication on Real Hydraulic System", in: SCYR 2011 : 11th Scientific Conference of Young Researchers of Faculty of Electrical Engineering and Informatics Technical University of Košice : proc. - Košice : FEI TU, 2011 S. 388-391. - ISBN 978-80-553-0644-5.

[8] Š. Jajčišin, A. Jadlovská, "Laboratory model of Hydraulic System Control" (in Slovak), in: Electroscope – online journal for Electrotechnics, year 2011, No. III. ISSN 1802-4564, Available on the internet: http://147.228.94.30/index.php?option=com_content&view=article&id=280:riadenie-laboratorneho-modelu-hydraulickeho-systemu-&catid=34:cislo-32011-&Itemid=48.

Selected Methods Used to Analyze Sentiment

[1]*Martina TARHANIČOVÁ (1^{st} year)*
Supervisor: [2]*Kristína MACHOVÁ*

[1,2]Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

[1]martina.tarhanicova@tuke.sk, [2]kristina.machova@tuke.sk

***Abstract*—Today's society is overflowed with information. World Wide Web offers a wide range of possibilities for web-users to express their opinions e.g. through blogs, forums, chats and social networks. It is a difficult task to find relevant information about a topic we are interested in without sacrificing a lot of time to read every available source of information to make assumptions. Therefore to resolve this dilemma and to make it easy for web-users there exists a field of study about opinion mining and sentiment analysis. In this paper, we define what sentiment analyses is, problems that have to be taken into consideration and traditional methods used to analyze text. In the end, we outline future research direction.**

***Keywords*—Sentiment analysis, opinion classification, opinion analysis, dynamic coefficient**

I. Introduction

Sentiment analysis represents a domain, which is a firm part of the field of the social web analysis. The social web can be considered as an upgrade of the classic web. The classic web can be illustrated with an idea of a world-wide billboard - anybody can publish some information piece or make it accessible for public inspection on the billboard (anybody who has necessary skills in web page creation - but considerably greater amount of web users have abilities only for reading this published information). On the other hand, the social web or web 2.0 reinforces social interactions among user and provides an opportunity for a great majority of web users to contribute to web content. It can be said that it increases the number of web content providers. Social interactions among users are enabled by communication within social nets, by the possibility to contribute to web discussions, and so on [1].

II. Sentiment analysis

Sentiment analysis is part of computational linguistics, natural language processing, and text mining. It involves different research tasks, such as:

- subjectivity detection (Wiebe et al., 1999; Pang and Lee, 2004)
- polarity classification (Pang et al., 2002; Turney, 2002)
- intensity classification (Wilson et al., 2009; Brooke, 2009)
- emotion identification (Chaumartin, 2007; Katz et al., 2007)

Subjectivity detection aims to discover subjective or neutral terms, phrases or sentences, and it is frequently used as a previous step in polarity and intensity classification with the aim of separating subjective information from objective one. Adjectives (beautiful) and adverbs (perfectly) are capable of expressing subjectivity remarkably well. On the other hand, we have to count with others world classes to also achieve high accuracy e.g., verb (destroy). This subjective words are embedded into dictionary together with their polarity.

Polarity classification attempts to classify texts into positive, negative or neutral. It forms the basis for determining polarity of text as whole. There are three degrees of polarity: positive (excellent), negative (poor) and neutral (average). Determining polarity of words is tightly connected with switching polarity problem. Switching polarity can be done by negation which is the reason for extending polarity of words to determine polarity of combination of words (taking into account the whole sentences or parts of the sentence).

The intensity classification task goes a step further and tries to identify different degrees of positivity and negativity, e.g., strongly-negative, negative, fair, positive, and strongly-positive. It can be described by numbers or by words. Numerical description is helpful in case of processing on the computer. Intensity of polarity can significantly change polarity of collocation e.g., surprisingly good, highly qualitative.

Finally, the emotion identification task seeks to identify the specific emotion (e.g., sadness, fear, etc.) that best reflects the meaning of the text. [2]

Another problem, which inhibit achieving rightness of sentiment analysis, is opinion spam. In recent years, we notice increased demand for opinion classification, but there is almost no attention to examine the credibility of opinions in reviews. Since there is no quality control, anybody has an opportunity to write whatever on Web thus lower the quality of reviews. The biggest problem coming out of this chaos is deceiving public opinion. These days people usually come to the web to checkup on products that they have the intention to buy. Such spam can promote product even it is the worst on the market. We can distinguish three types of opinion spam: the first is the misguided opinion, the second one is an opinion that does not relate directly to the studied entity - object and the last one is not real opinion, but only a text not relevant for opinion analysis [3].

III. Selected methods

A. Machine learning

1) Naive Bayes: The Naive Bayes algorithm is a widely used algorithm for document classification. There are two commonly used models (i.e., multinomial model and multivariate Bernoulli model) for text categorization [4]. Naive Bayes-based text categorization still tends to perform surprisingly well (Lewis, 1998); indeed, Domingos and Pazzani (1997)

show that Naive Bayes is optimal for certain problem classes with highly dependent features. On the other hand, more sophisticated algorithms might (and often do) yield better results [5].

2) SVM (Support Vector Machine): Support vector machines (SVMs) have been shown to be highly effective at traditional text categorization, generally outperforming Naive Bayes (Joachims, 1998). They are large-margin, rather than probabilistic, classifiers, in contrast to Naive Bayes and MaxEnt. In the two-category case, the basic idea behind the training procedure is to find a hyperplane, represented by vector *w* that not only separates the document vectors in one class from those in the other, but for which the separation, or margin, is as large as possible [5].

3) Maximum entropy: Maximum entropy classification (MaxEnt, or ME, for short) is an alternative technique which has proven effective in a number of natural language processing applications (Berger et al., 1996). Nigam et al. (1999) show that it sometimes, but not always, outperforms Naive Bayes at standard text classification [5].

TABLE I
COMPARISON OF PERFORMANCE USING DIFFERENT DICTIONARIES WITH SO-CAL

Dictionary	Overall percentage of correctness
Google-Full	62.98
Google-Basic	59.25
Maryland-Full-NoW	32.65
Maryland-Basic	58.16
GI-Full	68.02
GI-Basic	64.23
SentiWordNet-Full	65.02
SentiWordNet-Basic	61.47
Subjectivity-Full	72.04
Subjectivity-Basic	66.51
SO-CAL-Full	78.74
SO-CAL-Basic	66.04

TABLE II
EVALUATION OF SENTENCE USING DYNAMIC COEFFICIENT K

Sentence	Veľmi	sa	mi	páči	ten	telefón.
Evaluate	4	0	0	1	0	0

B. Semantic orientation

Opposite to machine learning, semantic orientation is a measure of subjectivity and opinion in text. It usually captures an evaluative factor (positive or negative) and potency or strength (degree to which the word, phrase, sentence or document in question are positive or negative) towards a subject topic, person, or idea (Osgood, Suci, and Tannenbaum 1957) [6].

1) Semantic Orientation - Calculator (So-CAL): In [6] was used lexicon-based approach to extract sentiment from text. SO-CAL begins with two assumptions: that individual words have what is referred to as prior polarity, that is a semantic orientation that is independent of context; and that said semantic orientation can be expressed as a numerical value.

SO-CAL is based on different dictionaries for adjectives, nouns, verbs and adverbs. As it is known from previous research, adjectives or adjective phrases are the primary source of subjective content in a document. The dictionary can be created in different ways: manually, using existing dictionaries such as the General Inquirer (Stone et al. 1966), or semi-automatically, making use of resources like WordNet (Hu and Liu 2004; Kim and Hovy 2004; Esuli and Sebastiani 2006). The dictionary may also be produced automatically via association, where the score for each new adjective is calculated using the frequency of the proximity of that adjective with respect to one or more seed words. Seed words are a small set of words with strong negative or positive associations, such as excellent or abysmal. In principle, a positive adjective should occur more frequently alongside the positive seed words, and thus would obtain a positive score, whereas negative adjectives would occur most often in the vicinity of negative seed words, thus obtaining a negative score. Dictionaries for nouns, verbs, and adverbs were hand-ranked using the same scale as adjectives dictionary (from -5 up to +5). Depending on polarity there are two categories to classify intensifiers. Amplifiers increase semantic intensity and downtoners decrease semantic intensity. Range of amplifiers and downtoners is given in percentage in between range -100% up to 100%. The simplest approach to the negation is simply to reverse polarity of a lexical item next to negator e.g., good (+3) into not good (-3). This may refer to as switch negation. Negation search in SO-CAL includes two options: Look backwards until a clause boundary marker is reached; or look backwards as long as the words/tags found are in a backward search skip list, with a different list for each part of speech [6]. This seems to work well in certain cases, but it fails in other. The key point is that if we consider e.g., excellent a +5 adjective negation of this word is not excellent, and this is far from atrocious a -5 adjective. In fact, not excellent is more positive than not good a -3 adjectives. Such an issue can be captured by implementing another method, polarity shift. Instead of changing just sign, value of negation is just recalculate with a fixed number 4. Thus, a +5 adjective becomes a +1 adjective and a -5 adjective become a -1 adjective. Polarity shifts seem to better reflect pragmatic reality of negation.

She is not terrific (5-4) but not terrible (-5+4) either. (1)

SO-CAL [6] is compared with others available dictionaries to see how good it is. In comparison were tested two options: full uses all default SO-CAL features and basic which was just sum of the SO words in relevant texts. Dictionary chosen to be compared with SO-CAL were Google-generated PMIbased dictionary, "Maryland" dictionary, General Inquirer and SentiWordNet dictionary. It can be concluded that SO-CAL results achieve slightly better results than others dictionaries. It identifies in overall correctly 78.74% Table.I.

2) Dynamic coefficient K: Coefficient K [7] is needed in order to evaluate sentiment successful in every sentence thus there is a need to set coefficient automatically which is able to adapt and choose the right length of words combination. This step is unavoidable to accomplish successful and accurate result. Subjective texts often contain words which change the meaning of words.

Sentence [1] Table. II would be considered as weak positive

[1]Experiments were done on Slovak language sentences therefore we provide examples written in Slovak.

TABLE III
EVALUATION OF SENTENCE USING DYNAMIC COEFFICIENT K

Method	Accuracy
1. method	80%
2. method	84%
3. method	82%

because of word ”páči”. In case of use coefficient K these cases could happen:

K=1. Program would score word ”veľmi” as 0 and so on till word ”páči” which would have value of 1. So final value of the whole sentence is 1.

K=2. Program would score the whole sentence the same as in case K=1. Only difference would be that, in this case, two words would be evaluated. It would begin with ”veľmi sa” which value is 0 and so on until ”mi páči” which would be again 1.

To achieve desired result, we have to set dynamic coefficient to value at least 4. Such case means that word with rising intensity would make a block of words with three following words ”Veľmi sa mi páči” with value 2.

Options to determining dynamic coefficient K:

Average length of sentences: To calculate dynamic coefficient, firstly every word in the sentence is calculated. Subsequently an average value is computed. Result value is used for every sentence to determine polarity.

Length of sentence / 2 (grounded up): Dynamic coefficient is set for every sentence independently.

Combination of first two approaches: Value of dynamic coefficient would be set as an average from first and the second method.

From experiments, we can say that there is a small difference in results while using above mentioned methods Table. III. Dynamic coefficient computed from the length of sentences came out as the worst method. Reviews of products choosen for testing purposes were from http://recenzie.sme.sk.

IV. CONCLUSION

As we can see from work done so far there are still areas that need to be further worked on. Results obtained using methods mentioned above are not the worst, but on the other hand, they are not sufficient. We propose to improve dynamic coefficient K by obtaining it dynamically depending on text; next we would like to incorporate coefficient of intensification and use besides switch negation also shift negation. More advanced level of research could be to determine if the text contains sarcasm or irony and then decide if the author of text has either positive or negative opinion. [8]

ACKNOWLEDGMENT

Research supported by the ”Center of Competence of knowledge technologies for product system innovation in the industry and service”, with ITMS project number: 26220220155 for years 2012-2015.

REFERENCES

[1] K. Machová, “Opinion analysis from the social web contributions,” in *Proceedings of the Third international conference on Computational collective intelligence: technologies and applications - Volume Part I*, ser. ICCCI’11. Berlin, Heidelberg: Springer-Verlag, 2011, pp. 356–365. [Online]. Available: http://dl.acm.org/citation.cfm?id=2041845.2041885

[2] J. C. de Albornoz, L. Plaza, and P. Gervs, “Sentisense: An easily scalable concept-based affective lexicon for sentiment analysis,” in *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC’12)*, N. C. C. Chair), K. Choukri, T. Declerck, M. U. Doan, B. Maegaard, J. Mariani, J. Odijk, and S. Piperidis, Eds. Istanbul, Turkey: European Language Resources Association (ELRA), may 2012.

[3] P. Szabo and K. Machova, “Various approaches to the opinion classification problems solving,” in *Applied Machine Intelligence and Informatics (SAMI), 2012 IEEE 10th International Symposium on*, Jan., pp. 59–62.

[4] S. Tan, X. Cheng, Y. Wang, and H. Xu, “Adapting naive bayes to domain adaptation for sentiment analysis,” in *Advances in Information Retrieval*, ser. Lecture Notes in Computer Science, M. Boughanem, C. Berrut, J. Mothe, and C. Soule-Dupuy, Eds. Springer Berlin Heidelberg, 2009, vol. 5478, pp. 337–349. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-00958-7_31

[5] B. Pang, L. Lee, and S. Vaithyanathan, “Thumbs up?: sentiment classification using machine learning techniques,” in *Proceedings of the ACL-02 conference on Empirical methods in natural language processing - Volume 10*, ser. EMNLP ’02. Stroudsburg, PA, USA: Association for Computational Linguistics, 2002, pp. 79–86. [Online]. Available: http://dx.doi.org/10.3115/1118693.1118704

[6] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, “Lexicon-Based Methods for Sentiment Analysis,” *Computational Linguistics*, vol. 37, no. September 2010, pp. 1–41, 2011. [Online]. Available: http://www.sfu.ca/~mtaboada/docs/Taboada_etal_SO-CAL.pdf

[7] M. Danko, “Dynamic coefficient in opinion acquisition from web discussions forums,” Apr. 2012.

[8] K. Shibahara, N. Inui, and Y. Kotani, “Adaptive strategies of mtd-f for actual games.” in *CIG*. IEEE, 2005. [Online]. Available: http://dblp.uni-trier.de/db/conf/cig/cig2005.html#ShibaharaIK05

Semantic Search Web Agent

[1]*Stanislav DVORŠČÁK (4^{st} year)*
Supervisor: [2]*Kristína Machová*

[1,2]Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

[1]stanislav-dvorscak@solumiss.eu, [2]kristina.machova@tuke.sk

***Abstract*—Article describes realization of Semantic Search Web Agent over Wikipedia dumps. The solution is realized over technologies like Evolution algorithms, Markov Models, Semantic Networks and Time Delay Neural Networks. There are waited properties of solutions like assistance for working with contextual respectively hidden information, misspelling support and aggregated results over searched documents.**

***Keywords*—Wikipedia, Semantic Web Agent, Time Delay Neural Network**

I. Introduction

The internet is definitely inseparable part of our life. This space contains a lot of accessible information pieces, which are readable for people, but they are unreadable for machines. To be able to use these information, it is necessary to have search engines. Search engines are based on several methods with support for several interactions. They can be split up over two common principles:

- Key search, which is based on using keywords, words or phrases. It is realized by searching in a index prepared during indexation. There are using a lot of technologies, which are less or more sophisticated like tokenizers, filters and so on. They are responsible to transform input documents into terms. And these terms are output of indexation process and they are base for searching. The problem is that context of the information is not used by them - like relations, knowledge, etc.
- Semantic search, which is based on semantic of searched information. In dependent on used architectures and ways of searching, they can be less or more intelligent with support for using contextual information, be able to use synonyms in general way, to find hidden but predictable information, and etc.

One of targets of the W3C consortium is also a semantic web. It describes a style how to create web sites to be in the form, which is good readable for people, but also for machines. If a machine wants to understand the needed information and extract it from a site, it has to know how to find, infer or derive this information. It is feasible by using state-of-the-art technologies - semantic technologies, for example metadata, ontology, monotonic and nonmonotonic logic [1]. Nonmonotonic logic are better for processing of the web information, because the web contains contradictory information and nonmonotonic logic can process the contradictory information and can be used for reasoning of new information [2]. But the most problem is that suggested technologies need human driven transformation of information, into the technologies like OWL2, RDF, rules etc.

We are implementing Semantic Web Agent over articles obtained from Wikipedia dumps. Support for misspelling tolerance, hidden information extraction, contextual information support are core requirements for implementation. In other words to process semantic information inherited from plain text. But it is plain text, which is not preprocessed to be structured that can be easily readable by machines.

II. Time Delay Neural Network/Semantic Network

The core of implementation is done by semantic network, which is customized in several directions.

Common semantic networks are defined as relations between concepts. It is knowledge representation. Where everything is represented in a form of graph. In dependent on concrete implementation, it can support reasoning, it can be organized into the taxonomic hierarchy and etc. The crucial problem of a lot of implementations is, that they are build by human with extensions done by machines. It is parallel representation.

But what about web and relation to mentioned semantic network? We want to have semantic web agent, which is done over information obtained from web. And this agent have to operate over semantic unstructured information.

Semantic network is not enough for processes representation. It has to be extended by time. The time is crucial dimension of our world. The extension by time has two positive impacts. First one - we are able to represent processes, which has meaning only in a context of time. And second one - rules in form of logic are used for processes representation. But if a original space of representation is extended by time, we can have support for non-monotonic logic. Because information can have relation not only to source and several metadata, but also to point of time.

On the figure 1. an overview of our semantic network is provided. It is digraph of symbols as vertices and time delay relations as edges.

The difference of our semantic network is in the relations. In the common semantic network relations have meaning. And they hold information. For us, this kind of information is represented again as symbols. And relations hold only information about time delay.

A formal semantic network formula is: $N = (S, C)$, where S is the set of symbols, objects (nodes) and the C is the set of connections (edges). Because, it is possible to have multiple connections to the same symbol, the aggregation function have to be defined - f_a.

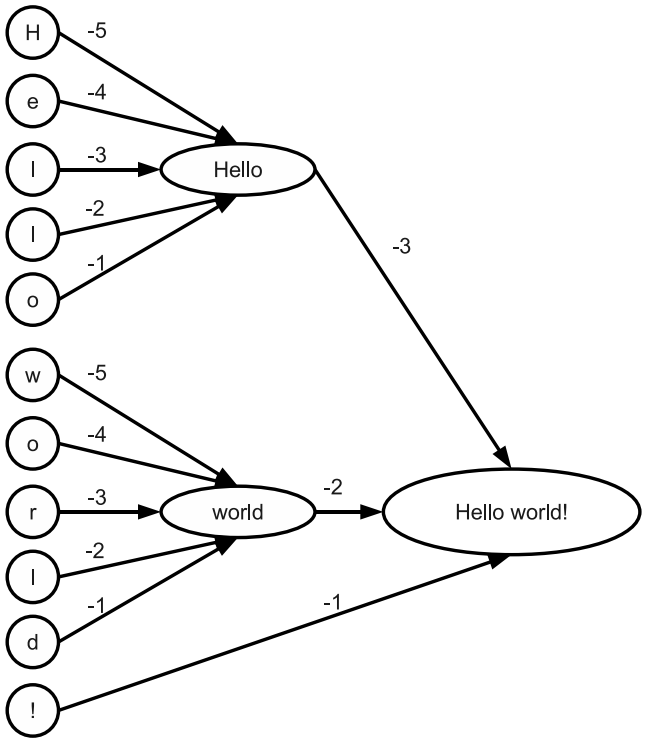


Fig. 1. Semantic Network

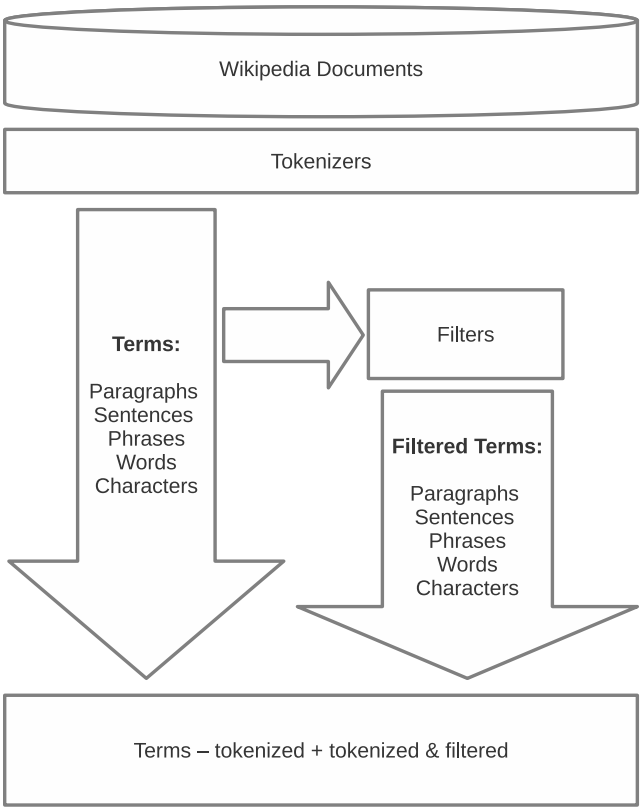


Fig. 2. Documents to terms transformation

Symbol - $S =< id, f_a, x >$

id: identity of symbol

f_a: aggregation function AND(weighted distance, ...), OR(max)

x: strength of signal $< -1, 1 >$, -1 invalid, 0 - N/A, 1 - valid

Connection - $C =< s_{input}, s_{output}, t_{delay}, e_{inv} >$

s_{input}: input symbol

s_{output}: output symbol

$t_{delay} = 1...n$: time delay of signal transition, where $t_{delay} > 0$

e_{inv}: flag - validity/invalidity

A. Input layer

Input layer represents a way how to fire symbols, respectively to start inference. It is also start point of searching where inference is used for semantic network building.

Each symbol has identity, which is unique over whole semantic network. The symbols of input are represented by terms. Terms are obtained by tokenizers and filters. There is difference in front of tokenizers and filters used by keyword search engines.

In our case a purpose of tokenizers and filters is splitting up documents into the next terms:

1) Paragraphs - Purpose to have paragraphs is to be able to build an overview of founded information, and also to be able to work on high level context information.
2) Sentences - provide support for overview of founded information and way how to transform founded symbols back to sentences. They are also cross references between sentences.
3) Phrases - provide support for overview of founded information and they are important for quick reasoning.
4) Words - provide support for quick reasoning.
5) Characters - they slow down inference process, but they are important in case of misspelling support.

Filters are used over all previously mentioned terms, but as extension of terms. Not replacement, as they are obviously used by keyword search engines. They are able to remove diacritics, to make lowercase respectively upper case, and etc.

For overview see figure 2.

B. Middle layer

Middle layer contains symbols, which have no meaning, respectively their meaning is hidden. They provide support similar as hidden states of hidden Markov model [3].

Symbols of these layer are generated and destroyed by several strategies, which are based on probability of symbols fired inside a window. For example, two symbols inside the window will create additional hidden symbol with relation to both of them.

The purpose is, if there exists any semantic relation between symbols, there exists also a probability of the same sequence of fired symbols. These relation is not visible for us, but it is there and it has influence on searching.

C. Output layer

When we are speaking about semantic search, it is more about asking for information as documents searching. Documents have reason for references - and self proof check. But they are not searched information.

For us, searching is represented as inference of semantic network. During inference, relevant symbols are founded. And the question is how to return from these symbols back to information, which will be presented as answer. Returned information are currently represented as structure of:

1) Documents - most relevant documents, which are used as search result, respectively references.
2) Words - most relevant words, which can be used as the extension of query.
3) Overview of founded information, it is attempt to build from founded symbols paragraphs, sentences, words, characters an abstract of searched information. It is kind of reversed process. It is experimental functionality.

We are in belief that to be able to solve third mentioned result information, it is necessary to have context of searching. And it can be influenced only as part of inference. But for success, we will consider also to solve first two points of result. And third point is about thinking, how to use fired symbols.

III. Indexing/Learning

There are articles of Wikipedia + external knowledges on the input of indexation process. Each term - article, paragraph, sentence, phrase, word and character get an input symbol. It is new or existing symbol based on identity of term. Articles, paragraphs, sentences have identity defined in the relation to article. It means, they are consider to be independent between articles. Phrases, words has identity based on equality - occurrence of same sequence of characters. Characters are consider to be terminal symbols and identity is based on character code.

Learning is similar to searching. Input symbols are inferenced, and based on statistics and probability hidden symbols together with relations will be created or destroyed. Opposite to searching, on the output are supplied the same sequence of symbols. And again based on statistics and probability there are created relations to hidden symbols. Input symbols are related to output symbols, but they are not the same.

IV. Searching

There are characters on the input, they are tokenized to terms in the same way as during indexation. Appropriate symbols are founded or created for terms. There were created relations during indexation process, which are used during searching to fire next symbols. And these symbols fire additional symbols. When output symbols are fired, they are consider to be output of searching. And some kinds of them can be used also as input of searching, it means suggestion for query extension.

V. Domains

Symbols and relations are logically grouped into the domains. Domain is consider to be also a symbol, which is fired together with the input symbols during indexation and searching. Domain are predefined by system - e.g. languages, source, user. But they are selectable, it means they are used based on the certain conditions. They also protect the system, because relations have meaning in context of domains, and has no influence on whole system. Relation is fired only in context of domains. We are able to use feedback during searching inside semantic network, but with no risk that semantic network will be corrupted or that there will be big noise inside semantic network.

VI. Misspelling Support

The implementation has misspelling support, which was currently tested and verified. But what is necessary to be mentioned, that this implementation has no influence to prefixes and suffixes of words. Which are important in case of languages, where these derivations are responsible for meaning.

Keyword search engines are using terms as a base element of searching. Input documents are transformed into terms during indexation by the tokenizers and filters. Tokenizers are responsible for spliting up documents into the mentioned terms and filters are responsible for removing or suppression of information, which are too specific. For imagination it is removing prefixes, suffixes, diacritics, verbal forms, prepositions, etc. But this one can not be realized in context of semantic, because they are responsible for meaning.

Because of this one, we do not remove these information from created semantic network, instead of removing, we extend original information by their derivations directly inside semantic network. After that, we are able to benefit from original information and also to be able to do cross searching over similar information. It has positive effect to semantic information and also redundancy done by derivations has positive effect for case of misspelling support and similar meaning.

VII. Markov Model Optimization

The performance respectively inference has to be optimized because of the implementation mentioned in previous sections.

Dynamic programing was chosen for implementation and as optimization was chosen markov model, which drives inference mechanism. It can be compared to evolution algorithms, where probability of markov model is used as a selective pressure.

For imagination, we have semantic network, where symbols represent states of markov models. And probability is determined by transitions between symbols. [3]

Evolution algorithms can be imagined as individuals generation, it means to generate path of inference and selective pressure of this individual respectively inference path is driven by mentioned probability.

VIII. Conclusion

The work is still in the progress, but we believe that we are close to the end. Ideas were verified on several prototypes and have success on several domains. Now we are working on whole Wikipedia dump with idea to provide semantic search over this source for public.

References

[1] G. Antoniou and F. van Harmelen, *A Semantic Web Primer*. Massachusetts Institute of Technology, 2004.
[2] G. Antoniou, *Nonmonotonic Reasoning*. The MIT Press, April 1997.
[3] A. Krokavec, D. – Filasová, *Optimal Stochastic Systems*. Elfa, 2002.

Semi-automatic annotation tool for aspect-based sentiment analysis

[1]*Miroslav SMATANA (3rd yearBc.),* [2]*Peter Koncz(3rd year)*
Supervisor: [3]*Ján PARALIČ*

[1,2,3]Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

[1]miroslav.smatana@student.tuke.sk, [2]peter.koncz@tuke.sk, [3]jan.paralic@tuke.sk

***Abstract*— Sentiment analysis, allowing automated quantification of textual evaluations, is becoming an important research area of recent years. Aspect-based sentiment analysis is an extension of classical sentiment analysis, which focuses on different features or aspects of the evaluated object. However, the creation and evaluation of the methods of sentiment analysis usually depends on manually annotated corpora, which creation is a time consuming task. In this paper we present a semi-automatic annotation tool for aspect-based sentiment analysis, which improves the efficiency of annotation by offering suggestions for annotations, which can be subsequently accepted or corrected by the user. The annotation tool was evaluated on the domain of hotel reviews with promising results.**

***Keywords*— sentiment analysis, semi-automatic annotation, text mining.**

I. Introduction

In last few years with the rising number of discussion forums, social networks, comment systems and other social web applications became more and more common that people post their opinions on internet. They do so especially when they buy some new products or services. These reviews created by users are useful information source for new customers. Amazon[1] was one of the first companies which allowed customers to evaluate their products on their website in 1995. Nowadays Amazon together with Facebook, Twitter and Google+ are essential sources of opinions and attitudes. The Pew internet research [1] has shown that 58% of respondents searching for information about products on internet and 24% of respondents comment or write reviews about products. The huge amount of user generated reviews on internet makes it really hard to manually browse them. One possibility how to deal with this problem is to use methods of sentiment analysis. Sentiment analysis enables the utilization of unstructured textual evaluations of products and services by their quantification [2]. To achieve a good precision of these methods it is necessary to build large corpora annotated especially for the needs of sentiment analysis. Existing annotation tools like GATE[2] (General Architecture for Text Engineering) aren't dedicated for this purpose and the annotation can be time consuming. The aim of our tool is to make annotation related to the aspect-based sentiment analysis more efficient by providing suggestions of annotations, which can be subsequently accepted or corrected by the user. In the following sections related works are first described. The proposed solution is then described followed by its experimental verification. On the basis of obtained results the findings are summarized in the conclusion and further research goals are formulated.

[1]http://amazon.com
[2]http://gate.ac.uk

II. Related Works

Methods of sentiment analysis can be divided into two groups, namely endogenous and exogenous [2]. These groups differ by data which they use. While endogenous methods estimate the sentiment only on the basis of the annotated training sample and the used learning method, exogenous methods use also external knowledge like emotional word dictionaries or manually created rules. One of the biggest constraints of endogenous methods is their dependency on huge amount of annotated training documents. To facilitate their creation it can be utilized annotation tools with some level of automation, usually in form of annotation suggestions. Dill et al. [3] proposed an automatic semantic annotation tool called SemTag. It detects occurrences of entities and then it uses the so called Taxonomy Based Disambiguation (TBD) algorithm. Kiryakov et al. [4] described in their work a semantic annotation platform called KIM (Knowledge and Information Management). KIM uses ontology based annotation, where named entities are annotated and linked to the concepts of the ontology. KIM uses GATE for information extraction and for content and annotations management. In the experiment described in [4] gazetteers, shallow analysis of the text and pattern matching grammars were used for information extraction, however the proposed architecture of KIM allows integration of other GATE pipelines. Song et al. [5] discuss in their work a tool called Semantator. Semantator annotates parts of biomedical texts according to ontologies and this annotation can be confirmed or corrected by users. For supporting of the curation of biomedical texts Rak et al. [6] developed a tool called Argo based on UIMA (Unstructured Information Management Architecture) framework. There are also annotation tools which provide general platforms for language processing tools. Such platforms are for example the mentioned GATE and UIMA. An overview of annotation tools with automated creation of annotations can be found also in the work of Kiyavitskaya et al. [7].

Our annotation tool differs from other annotation tools by its specialization to the aspect-based sentiment analysis. It uses mainly endogenous method based on machine learning both for the aspect identification as well as sentiment analysis.

III. PROPOSED SOLUTION

The proposed solution, which architecture is depicted on the figure 1, is a semi-automatic annotation tool. The main parts of its architecture are presented in the following subsections. The tool consists of preprocessing module, classification module and annotation module and is created as an extension of the tool called Luwak[3], which is a more general semi-automatic annotation tool. It is built on Eclipse platform and uses embedded version of GATE. The main function of Luwak is the support of the Eclipse plug-in architecture for customization of user interface to task specific needs. Knowledge engineer is able to customize text preprocessing steps, conceptual model for extraction, graphical user interface for manual annotation and automatic extraction algorithms.

A. Domain model creation

After initial data gathering process it is necessary to build an appropriate domain model. For the purpose of aspect-based sentiment analysis realized on the domain of hotel reviews we selected the following attributes of hotels: cleanness, comfort, location, services, personnel, price-quality ratio. These attributes correspond with the aspects of evaluations.

B. Preprocessing

To achieve the required accuracy of suggestions appropriate preprocessing of texts is needed. In our tool we use preprocessing modules from ANNIE[4] which is GATE's information extraction system. We decided to use three modules namely: sentence splitter, tokenizer and stemmer. Sentence splitter divides the entire text to individual sentences which are then processed. Tokenizer divides text to words or phrases which we call tokens. Stemmer cuts the suffixes from tokens. In our tool we used Porter stemmer. Sentences are divided into 1-3 grams, which are n words going in sequence. For example if we have a sentence "A great hotel in good location, close to shops, entertainment, bars and transport links." 2-grams are "great hotel", "hotel in", "in good", "good location" etc. These n-grams create a large feature space. We used feature selection on the basis of the information gain to reduce the feature space dimensionality. Reduction of the feature space size can improve the classification accuracy and reduce the model creation time. Within the existing feature selection methods information gain has been shown as one of the best [8] and its effectiveness was confirmed also in the context of sentiment analysis [2]. According to Yang and Pedersen [8] information gain measures the number of bits of information obtained for category prediction by knowing the presence or absence of a feature in the document. This can be expressed by the following formula.

$$G(t) = -\sum_{i=1}^{m} P_r(c_i)\log P_r(c_i) + P_r(t)\sum_{i=1}^{m} P_r(c_i|t)\log P_r(c_i|t) + P_r(\bar{t})\sum_{i=1}^{m} P_r(c_i|\bar{t})\log P_r(c_i|\bar{t})$$

In which t is a feature and c_i is a category. The preprocessing stage was realized once on the whole dataset. However the feature selection belongs to the preprocessing stage in each retraining phase it was necessary to compute the actual values of features information gain, so it is mentioned under the sentence classification.

C. Classification

Sentence classifier is a module which is designed for classification of sentences according to the mentioned aspect of evaluation as well as the polarity and intensity of sentiment. It is necessary to classify each sentence of document separately, because each sentence can represent different aspects and sentiments. Sentiment is annotated on a five point scale from very bad to very good. We used Naïve Bayes classifier both for the aspect classification as well as sentiment classification. The classification model is periodically retrained on annotated sentences.

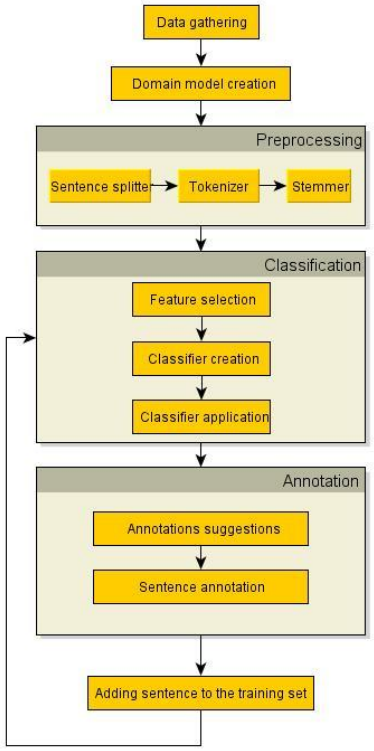


Figure 1. Architecture of annotation tool

D. Annotation

In this module we do annotation of sentences. Inputs to this module are values from naïve Bayes classifier which are evaluated according to rules. More aspects can be assigned to one sentence if they have similar probability. Because the numerator and denominator are very small numbers, we use logarithms to avoid underflow. We evaluate whether the probabilities of classes differ significantly from the probabilities of other classes. Otherwise annotation tool marks the aspect with the highest probability.

On figure 2 it is shown the user interface of our annotation tool. At the left pane you can see the list of documents. Documents are displayed in the central part where the currently annotated sentence is highlighted. At the right pane the annotation suggestions are displayed which can be confirmed or corrected by the user.

[3]http://sourceforge.net/projects/luwak/
[4]http://gate.ac.uk/sale/tao/splitch6.html#chap:annie

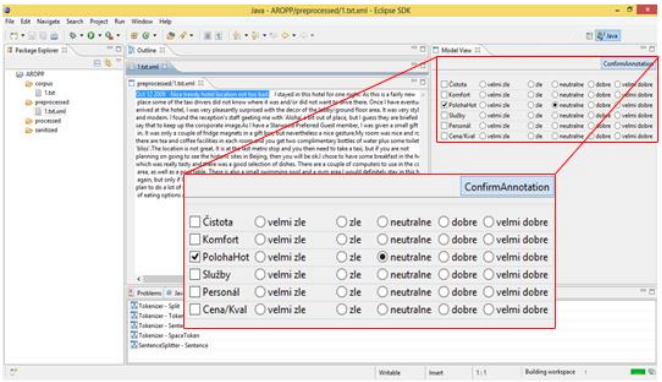


Figure 2. User interface of the annotation tool

The output is an XML file containing the annotations for each of annotated sentences in a format compatible with GATE. After the annotation of given sentence it is added to the training corpus, feature selection is realized and classifiers are retrained. For each aspect a separate sentiment classifier is trained.

IV. Experiment

A. Dataset

We tested our annotation tool on dataset from website KavitaGanesan[5]. On this website you can find over 250000 reviews about hotels from different countries. In our experiment we performed annotation of 250 sentences. These sentences were annotated by annotator and the suggestions as well as user annotations were logged.

B. Results

For the evaluation of the suggestions we used f-measure which is a harmonic mean of precision and recall. It was computed regardless and regarding to particular aspects. In figure 3 it is shown the relation between the number of annotated sentences and the value of f-measure.

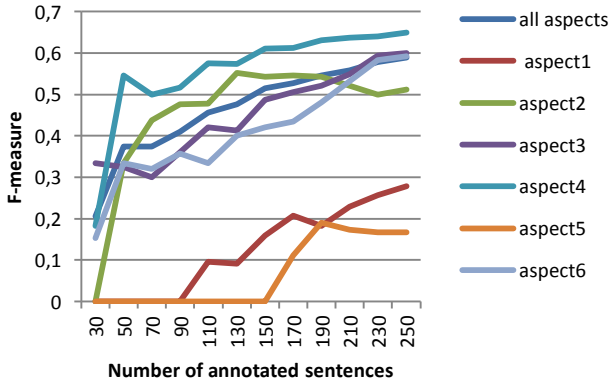


Figure 3. Relation between the number of annotated sentences and the value of f-measure

After the annotation of the whole set of sentences was the f-measure for all aspects more than 58%. Some aspects occurred in sentences rarely (for example aspect 1) so they brought errors to the results. Annotation tool was able to annotate sentences correctly although they contain more than one aspect. From the figure 3 we can see the rising tendency of the value of f-measure.

For the evaluation of sentiment analysis model we used relative standard error which considers numerical values and classification precision which considers nominal values. The achieved relative standard error was 1.77 and the precision was 54%.

V. Conclusion

In this paper we described a semi-automatic annotation tool which could be used to speed-up annotations for the needs of aspect-based sentiment analysis. The achieved results show that the proposed pipeline of documents processing can be used to get accurate annotation suggestions. However, more detailed evaluation on bigger dataset should be done to confirm the increase of annotation speed-up according to classical annotation tools. We will also try to increase the efficiency of the proposed solution by leveraging active learning methods.

Acknowledgment

This work was partially supported by the Scientific Grant Agency of the Ministry of Education, Science, Research and Sport of the Slovak Republic and the Slovak Academy of Sciences under grant No. 1/1147/12 and partially by the Slovak Research and Development Agency under the contract No. APVV-0208-10.

References

[1] J. Jansen, "Online Product Research: 58% of Americans have researched a product or service online," 2010. [Online]. Available: http://pewinternet.org/Reports/2010/Online-Product-Research.aspx.

[2] P. Koncz and J. Paralic, "An approach to feature selection for sentiment analysis," in *2011 15th IEEE International Conference on Intelligent Engineering Systems*, 2011, vol. 15, no. June, pp. 357–362.

[3] S. Dill, N. Eiron, D. Gibson, D. Gruhl, R. Guha, A. Jhingran, T. Kanungo, K. S. McCurley, S. Rajagopalan, A. Tomkins, J. A. Tomlin, and J. Y. Zien, "A case for automated large-scale semantic annotation," *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 1, no. 1, pp. 115–132, Dec. 2003.

[4] A. Kiryakov, B. Popov, I. Terziev, D. Manov, and D. Ognyanoff, "Semantic annotation, indexing, and retrieval," *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 2, no. 1, pp. 49–79, Dec. 2004.

[5] D. Song, C. G. Chute, and C. Tao, "Semantator: annotating clinical narratives with semantic web ontologies," *AMIA Summits on Translational Science proceedings AMIA Summit on Translational Science*, vol. 2012, pp. 20–29, 2012.

[6] R. Rak, A. Rowley, W. Black, and S. Ananiadou, "Argo: an integrative, interactive, text mining-based workbench supporting curation," *Database*, vol. 2012, no. 0, p. bas010, 2012.

[7] N. Kiyavitskaya, N. Zeni, J. R. Cordy, L. Mich, and J. Mylopoulos, "Cerno: Light-weight tool support for semantic annotation of textual documents," *Data & Knowledge Engineering*, vol. 68, no. 12, pp. 1470–1492, Dec. 2009.

[8] Y. Yang and J. O. Pedersen, "A Comparative Study on Feature Selection in Text Categorization," *Proceedings of the Fourteenth International Conference on Machine Learning*. Morgan Kaufmann Publishers Inc., pp. 412–420, 1997.

[5]http://kavita-ganesan.com/entity-ranking-data#HotelReviews

Simulation of Empathy in Machines Incorporating the Subjective Nature of Human Emotions

[1]*Mária VIRČÍKOVÁ (3rd year),* [2]*Filip JERGA*
Supervisor:[3]*Peter SINČÁK*

[1,2,3]Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

[1]maria.vircikova@tuke.sk, [2]filip.jerga@student.tuke.sk, [3]peter.sincak@tuke.sk

Abstract— **In this paper, we present an interactive human-robot system with elements of reflecting emotions, inspired by the findings from the neurosciences about special cells called mirror neurons, believed to be the foundation of human empathy. Moreover, recent research in robotics has shown that the ability of machines express empathic emotions enhances human-machine interaction. We cope with the subjectivity of the stimuli used to elicit emotions for each of the users proposing a personalization of the robotic character based on subjective preferences of users during a human-machine dialog, where the most common used words during the conversation are assigned with a subjective emotional affect, performed by body-based expressions of the robot. The robotic expressions of empathic emotions enhance users' satisfaction, engagement, perception of robots, and performance in task achievement. We believe that the system allows to communicate with human beings in natural way and thus to participate in human social activities.**

***Keywords*—affective computing, human-robot interaction, subjective computing.**

I. INTRODUCTION

"Machines are becoming devastatingly capable of things like killing. Those machines have no place for empathy. There's billions of dollars being spent on that. Character robotics could plant the seed for robots that actually have empathy," – these are the words of David Hanson in his famous TED talk[1].

Machines cannot feel and express empathy. However, it is possible to build robots that appear to show empathy, which is commonly understood [37] as the capacity to "put your-self in someone else's shoes to understand his/her emotions." There is a convergence between cognitive models of imitation, constructs derived from social psychology studies on mimicry and empathy, and recent empirical findings from the neurosciences, e.g. [3][4][5][6][7][8][9][10] . According to [6], empathy research indicates that it is made possible by a special group of nerve cells called mirror neurons, at various locations inside the brain. Mirror neuron activity helps people understand actions and intentions of others and is also involved in understanding emotions. Empathy may be facilitated through a process of automatic mapping between self and other. Perception of the actions and emotions of others activates areas in our own brain that typically respond when we experience those same actions and emotions.

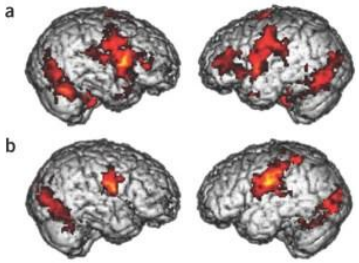


Fig. 1. In study by [7] , test subjects were asked to imitate facial expressions of others. fMRI scans revealed less brain activity in regions associated with mirror neurons in individuals with autism spectrum disorder (b) than in the control group (a).

The global challenge of our research is to design socially engaging robots and interactive technologies that provide people with long-term social and emotional support – in general to help people live healthier lives, connect with others, and learn better. In this paper, we present a system which reacts to the stimuli subjectively, depending on the user's preferences. For the implementation of the emotional conversation we use the SDK kit created by Microsoft [30], concretely its program for speech recognition. We enrich the conversation between the human and the machine, a humanoid robot, with the expressions of personalized emotions.

Fig. 2. Hanson's android reflecting his emotions as a mirror system that can be perceived a tip-off on artificial empathy. The artificial head reacts to the facial and verbal expressions of the human by

This paper is organized as follows:

Section II presents the state of the art of related research focused on the field of social robotics. Next section III analyzes the term of emotion together with the research on subjective computing, considered as a new stage of computer science. Section IV introduces our approach to build a system that reflects the human emotional affect during the communication and in this way; the system seems more believable and natural for the human partner.

II. Empathy in Human-Robot Interaction

The current research beyond the hypothesis that giving machines the capability of expressing empathic emotions towards users demonstrates its great potential on improving the overall interaction, e.g. in [36][35][34][33][31][32].

Breazeal [29] investigates emerging robotics applications for domestic or entertainment purposes which are slowly introducing autonomous robots into society at large. She claims that a critical capability of such robots is their ability to interact with humans, and in particular, untrained users. She explores the phenomenon that people will intuitively interact with robots in a natural social manner provided the robot can perceive, interpret, and appropriately respond with familiar human social cues.

In the area of social robotics, Kanda et al. [27] has achieved to improve route guidance interactions with a robot by incorporating cooperative body movements (e.g. synchronization of arm movements), enhancing both reliability and sympathy.

Riek et al. [28] studied the effect of automatic head gesture mimicking with a chimpanzee robot. The robot would listen to participants while either mimicking all head gesture, only nodding or no mimicking, resulting in different levels of interaction satisfaction. This work extends the state of the art by explicitly evaluating facial expression mirroring in contrast to head, arm or body gestures.

Coming back to the roots of artificial intelligence, Turing by himself was fascinated by the notion of affective responses, such as joy, interest, and surprise, in human interaction with computers [2]. Turing's initial questions have led to more complex investigations into the ability of the computer to model human emotions, and to evoke emotional responses in its human user.

A wide spectrum of existing projects and applications (e.g. [11][12][13][15][16][17][18][19][20]) are trying to better understand human emotional behavior and develop a model according to their needs and expectations, implementing this model in machines that interact with people. Such projects assume that during the process of system migration to human society they will be considered beneficial and intuitive partners. How we believe future cooperation between robots and ourselves will look can be summed up in these words: machines fully adapting to man – that man no longer has to adapt his behavior to machines.

III. On the Subjectivity of Emotions

We believe that equipping machines with the ability to simulate empathy can drift towards intuitive mutual man-machine cooperation. Inspired by the psychological models of emotions, researchers in artificial intelligence and cognitive robotics have begun to recognize the utility of computational models of emotions for improving complex, interactive programs. Software agents may use emotions to facilitate the social interactions and communications between groups of agents and this way they can help in coordination of tasks, such as among cooperating robots. Moreover, synthetic characters can use a model of emotion to simulate and express emotional responses, which can effectively enhance their believability. Furthermore, emotions can be used to simulate personality traits in believable agents. Recently, however, cognitive neuroscience and related fields have demonstrated the inseparability of emotion from rational thought and normal human function [39] – as Minsky believes, "the issue is not whether intelligent machines can have emotions, but whether machines can ever be intelligent without them [14]."

Emotions comprise subjective experience and expressive behaviour, are motivators for actions, and will change according to the range of actions that the subject is able to take in a given situation [21]. On the other side, computer science and robotics use techniques which have been developed through objectivism. The domain of human-robot interaction is different-machines are not only evaluated by objective measures, but also subjectively. This phenomena as a specific research phase represents a new stage of computing based on the subjectivity in the human perceptual process.

Suzuki [22] from University of Tsukuba, introduced an approach of Subjective Computing (SC) in 1995, tightly related to cognitive science and including an integrated physical robot system for investigating varieties of mechanisms relevant to embodied. A modelling taking insight from the human subjectivity and individual preference is the main issue of this research. SC includes the terms of individual emotional resonance, comfort and satisfaction. It differs from a conventional evaluation based on objectivity and logic used in physics or mathematics. The subjectivity and individual preference should be treated as psycho-physiological interrelationship.

Harrel [23] at MIT Imagination, Computation, and Expression Laboratory, uses the term of Subjective computing systems for artificial intelligence and cognitive science-based computing systems for creative expression, cultural analysis, and social change.

We try to cope with the subjective experiences that influence the emotional affect of the meanings of the words during the verbal communication with the robot, described in section IV.

IV. Implementation

The goal of the research is to create user friendly interface, where a humanoid robot expresses emotions according to the word said by the user. It is difficult to define emotional affects for each person specifically, because every person has different express of emotion on different situations and circumstances. Thus it is very difficult to create system which is specified for wide spectrum of users. We developed a personalized system—a user can assign the type of the emotion to given word on the web interface, and then the chosen emotion will be implemented into the robot, based on the individual preferences of the user. The system can identify a speech command from user, which is followed by expressed emotion of the robot (for example, if the human assigns the emotional affect of "Joy" to the word "mother", the robot performs the expression of joy every time when he recognizes

the word "mother" during the communication with the human).

To identify the types of emotions a user may feel during human-machine interaction, we explore the work of Plutchik [25]. Plutchik's psychoevolutionary theory of emotions is one of the most influential classification approaches for general emotional responses. He considered there to be 8 primary emotions - anger, fear, sadness, disgust, surprise, anticipation, trust, joy. Plutchik's proposed that these basic emotions are biological primitive and have evolved in order to increase reproductive ability of animals.

The program is written in a system environment of Microsoft Visual Studio 2010 Professional using the .NET platform because robot NAO and Kinect sensor supports this programming language and theirs API are well documented. Kinect [30] is a motion sensing input device by Microsoft for the Xbox 360 video game console and windows PCs. One of key feature of Kinect for Windows is speech recognition. Kinect contains 4 microphones linearly arranged. The sensor improves the sound quality through noise suppression and acoustic echo cancellation. Nao is humanoid robot created by Aldebaran Robotics, French company. His height is 57cm and the advantage is his simple manipulation. The API supports programming in various languages, e.g. C#, Python, C++, among others. We use the program API provided by the company Microsoft, concretely Microsoft for Kinect SDK beta, Microsoft speech platform, which is used for recognition of speech commands and API Naoqi created by Aldebaran Robotics for programming robot Nao.

The basic logic of the system consists in the following steps:

- A web interface for the creation of a personalized dictionary of words was designed. Here the user assigns the emotional affect based on his preferences.
- The personalized database of the emotional affects for the most common words used in the conversation with the humanoid robot is used during the proximate interaction.
- The Kinect sensor captures the word, compares it to the words saved in the database and if the human uses a word that is situated in the dictionary, the robot expresses assigned emotion. The speech commands are written in the form of dictionary.
- This process runs in cycle from start until end of program. If application does not find connected Kinect to PC, a message, informing user that Kinect was not find appears and the application is executed.

According to the Plutchik's theory of emotions, eight types of basic emotions are created. As the humanoid platform Nao does not have face to create facial expressions, we implemented body-based expressions of emotions, by setting angles of the joints of the robot. Expressed emotions are based on technical capabilities of Nao and common human emotion expression.

Fig. 3. Human interacting with the humanoid robot.

Fig. 4. Robot expresses Fear.

```
case "Trust":
    LShoulderRollAngles = new List<float> { 0.157f };
    LShoulderPitchAngles = new List<float> { 0.478f };
    LElbowRollAngles = new List<float> { -1.445f };
    LElbowYawAngles = new List<float> { 0.21f };
    LWristYawAngles = new List<float> { 0.0f };
    RShoulderRollAngles = new List<float> { -0.23f };
```

Fig 5. Part of the code – Emotion of Trust

V. Discussion

Emotions and their expressions belong to the nature of every human, but are not so common in robots and computing systems. Systems capable of expressing emotions may have big impact not only on field of computing technology but psychology as well.

This paper introduces our approach of the interactive human – robot system with elements of emotions. The results of the work can be viewed also from a certain psychological point of view. The goal is not to create an application that serves only for entertainment, but also to receive some sort of feedback in the form of information and experiences from users. There are countless possibilities where a robot which can express empathy could help. Some people can open themselves and express emotions, tell their thoughts to machines rather than to humans. This implies that these systems can help many people. On the other hand, the system can be used for entertainment of people, for example in art, where robot could perform any theater play.

Now we need to conduct various experiments to prove whether mirroring emotional affects has a positive influence on the human-robot interaction. Giving emotional meaning to a word in conversation is one part of our greater project on human-robot empathy. The other parts include for example gesture recognition using neural networks, a learning framework for body-based expressions from human to robot and the emotional model, where the basic emotions are combined into complex emotional spectrum. Moreover, we develop a visual system for monitoring the interest of people in the interaction with the robot.

Future work will explore the contribution of the simulation of empathy in real environments, measuring and comparing the interest of human subjects in the interaction with and without the elements of the artificial empathy.

This work is part of a wider tendency to rethink the way that interfaces operate: "a shift toward a human-centered interaction architecture, away from a machine-centered architecture[38]." This shift recognizes the importance of human emotion and response in the design of interfaces and points up some changes in the human-machine relationship. For example, machines have begun to adapt to their users, rather than the other way round; the interface is becoming a responsive entity, rather than a passive portal.

REFERENCES

[1] D. Hanson, "Robots that show emotion," in *TED talks*, 2nd ed. vol. 3, J. Peters, Ed. New York: McGraw-Hill, 1964, pp. 15–64.

[2] S. Huma and K. Warwic, "Emotion in the Turing Test: A Downward Trend for Machines in Recent Loebner Prizes," in *Handbook of Research on Synthetic Emotions and Sociable Robotics: New Applications in Affective Computing and Artificial Intelligence.* IGI Global, 2009.

[3] M. Iacoboni , "Imitation, empathy and mirror neurons," in *Annu Rev Psychol*, 2009.

[4] P. Molenberghs, V. Halász , J.B. Mattingley, E.J. Vanman and R. Cunnington, "Seeing is believing: Neural mechanisms of action-perception are biased by team membership", in *Human Brain Mapping*, in press.

[5] P. Molenberghs, R. Cunnington, J. B. Mattingley, "Is the mirror neuron system involved in imitation? A short review and meta-analysis", in *Neuroscience and Biobehavioral Reviews*, 2009, 33, pp. 975-980.

[6] T. S. May, "Your Pain is my Pain-if you play a fair game", in *Terms of empathy*, 2006.

[7] M. Dapretto, M. S. Davies, J.H. Pfeifer, A. A. Scott, M. Sigman, S. Y. Bookheimer, M. Iacoboni , „Understanding emotions in others: mirror neuron dysfunction in children with autism spectrum disorders", in .*Nat Neurosci*. 2006 pp. 28-30.

[8] A. Lingnau, B. Gesierich, A. Caramazza, „ Asymmetric fMRI adaptation reveals no evidence for mirror neurons in humans", in *Proc Natl Acad Sci* , 2009, pp. 9925-30.

[9] J. M. Kilner, A. Neal, N. Weiskopf, K. J. Friston, C. D. Frith, „Evidence of mirror neurons in human inferior frontal gyrus", in *J Neurosci*, 2009, pp. 10153-9.

[10] I. Dinstein, C. Thomas , K. Humphreys, N. Minshew, M. Behrmann, D. J. Heeger, „Normal movement selectivity in autism", in *Neuron*. 2010, pp. 461-9.

[11] C. Breazeal , "Designing Sociable Robots", in *The MIT Press*, 2002.

[12] S. Saint-Aime, B. Le-Pedevic and D. Duhaut, " iGrace – Emotional Computational Model for EmI Companion", in *Advances of Human-Robot Interaction*, ISBN: 978-953-307-020-9, InTech, 2009.

[13] A. Beck, "Displaying realistic emotion for animated characters", 2007. http://www.di.uniba.it/intint/DC-ACII07/Beck.pdf.

[14] M. Minsky, "The Emotion Machine", in *Simon and Schuste*, 2006.

[15] M. S. El Nasr, J. Yen, T. Iorger, FLAME — Fuzzy Logic Adaptive Model of Emotions, in *Fuzzy Logic Adaptive Model of Emotions. Autonomous Agents and Multi-Agent Systems*, 3(3), 219-257, 2000. Mendeley

[16] N. Esau, L. Kleinjohann, B. Kleinjohann, "An Adaptable Fuzzy Emotion Model for Emotion Recognition", 2005.

[17] S. Grossberg, "Adaptive Resonance Theory", in *Technical Report CAS/CNS-2000-024*, Boston University Center for Adaptive Systems and Department of Cognitive and Neural Systems, September 2000.

[18] N. Mendiratta, "Building emotional machines", in *AI-Depot.com*

[19] M. Kipp, J.C. Martin, "Gesture and Emotion: Can basic gestural form features discriminate emotions?", in *Affective Computing and Intelligent Interaction and Workshops*, 2009. ACII 2009. 3rd International Conference on, ISBN 978-1-4244-4800-5, 2009.

[20] K. Suzuki, et al., „Intelligent agent system for human–robot interaction through artificial emotion", in *Proceedings of the IEEE SMC*, 1998.

[21] Ravaja, Niklas, et. al., „The Psychophysiology of Video Gaming: Phasic Emotional Responses to Game Events", in *Proceedings of DiGRA*, 2005,

[22] K. Suzuki, Center for Cybernics Research & Faculty of Engineering, Information and Systems , University of Tsukuba, http://i-www.iit.tsukuba.ac.jp/~kenji/

[23] D. F. Harrel, Digital Media Comparative Media Studies Program & Computer Science and Artificial Intelligence Laboratory, MIT, http://groups.csail.mit.edu/icelab/

[24] S. Turkle, "Whither Psychoanalysis in Computer Culture?", in *Psychoanalytic Psychology*, vol. 21, pp. 16-30, 2004.

[25] R. Plutchik, "Circumplex Models of Personality and Emotions", in *American Psychological Association*, 1997.

[26] Personal Robots Group, MIT, http://robotic.media.mit.edu/

[27] T. Kanda, M. Kamasima, M. Imai, T. Ono, D. Sakamoto, H. Ishiguro, and Y. Anzai, "A humanoid robot that pretends to listen to route guidance from a human," in *Autonomous Robots*, vol. 22, pp. 87–100, 2007.

[28] L. Riek, P. Paul, and P. Robinson, "When my robot smiles at me: Enabling human-robot rapport via real-time head gesture mimicry," in *Journal on Multimodal User Interfaces*, vol. 3, pp. 99–108, 2010.

[29] C. Breazeal, "Regulation and Entrainment in Human-Robot Interaction", in *The International Journal of Robotics Research*, 2002, vol. 21 no. pp. 10-11.

[30] Microsoft Kinect sensor Tutorial, http://www.microsoft.com/en-us/kinectforwindows/develop/tutorials/tutorialsdesc.aspx?tutorialid=Tutorial1

[31] H. Prendinger and M. Ishizuka, "The empathic companion: A character-based interface that addresses users' affective states", in *International Journal of Applied Artificial Intelligence*, 19, pp. 297–285, 2005.

[32] R. Picard and K. Liu, "Relative Subjective Count and Assessment of Interruptive Technologies Applied to Mobile Monitoring of Stress", in *Int. Journal of Human-Computer Studies*, 65, pp. 396–375, 2007.

[33] M. Ochs, C. Pelachaud, and D. Sadek, "Emotion elicitation in an empathic virtual dialog agent", in *Prooceedings of the Second European Cognitive Science Conference (EuroCogSci)*, 2007.

[34] M. Ochs, C. Pelachaud, and D. Sadek, "An Empathic Rational Dialog Agent", in *International Conference on Affective Computing and Intelligent Interaction (ACII)*, pp 338–379, Lisbon, Portugal, 2007.

[35] J. Klein, Y. Moon and R. Picard, "This computer responds to user frustration", in *Proceedings of the Conference on Human Factors in Computing Systems*, pp. 242–243. ACM Press, New York, 1997.

[36] H. Prendinger and M. Ishizuka, "The empathic companion: A character-based interface that addresses users' affective states", in *Int. Journal of Applied Artificial Intelligence*, 19, pp. 297–285, 2007.

[37] E. Pacherie, "L'empathie, chapter L'empathie et ses degr´es", pp 149–181. Odile Jacob, 2004.

[38] C. Lisetti and D. Schiano, „Automatic Facial Expression Interpretation: Where Human-Computer Interaction, Artificial Intelligence and Cognitive Science Intersect", in *Pragmatics and Cognition*, 8(1) pp. 185-235, 2000.

[39] A. Damasio, „Descartes' Error: Emotion, Reason and the Human Brain", Penguin, 1994.

Smart Teleoperation System Framework

[1]*Martin PAĽA (2nd year)*
Supervisor: [2]Peter SINČÁK

Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

[1]martin.pala@tuke.sk, [2]peter.sincak@tuke.sk

***Abstract*—This paper describes a novel approach to robot teleoperation system capable of task learning through teleoperation. The system is based on the task rule extraction from the data acquired by monitoring a human operator during the teleoperation process. The basic information about teleoperation systems in general, applications and current trends are provided in the first part of the paper. The second part of this paper is dedicated to the architecture and software implementation of the proposed teleoperation system. The last sections of this paper are dedicated to the conducted experiments and conclusion of the project and the possible improvements considered for the future work.**

***Keywords*—Teleoperation, learning through teleoperation, fuzzy, fuzzy cognitive maps**

I. Introduction

Teleoperation has very rich history dating to 1940s. Since then it led to many practical applications and brought a human ability to interact with remote environments using manmade machines, allowed us to manipulate toxic and very hazardous materials, explore the space and to perform many other things. Nowadays the teleoperation systems are not only considered to be used in hazardous environments, but are widely used all around the world in various applications from space to entertainment applications. To have an ability to control something remotely has a strong impact in the business sector, so there is a strong demand for teleoperation process improvement, and the main trend is to reduce the operator's share on the control process and increase the teleoperator's (virtual teleoperator's) share instead.

Nowadays, the robotics systems are becoming more and more complex and thus it is very difficult to operate them remotely using only standard teleoperation interfaces. This leads to a human operator fatigue problems, communication and synchronization problems between operators. Increasing the autonomy of the teleoperator leads not only to reducing the amount of work needed to be done by a human operator, but also gives us the opportunity to operate in places, where it was not possible earlier (time lags, unstable connectivity, etc...), and leads to reducing the mishaps of remotely controlled systems. There are many ways of building a knowledge base for the robotic systems to give them an ability to solve a task more or less autonomously. Unfortunately, there is almost always expert's presence needed during the programming of the robot. In addition, the most of the tasks programmed for these robotic systems are hard-coded, what is very time consuming. Moreover, in the most of the cases, it is not possible to program a robot to perform all its tasks that may be required in the real world applications.

II. Teleoperation and Current Challenges in Teleoperation Systems

In order to understand teleoperation, some of the basic terms and concepts need to be clarified. The concepts and terms provided in the following words appear in the well known articles related to teleoperation [1, 2, 3, 4, 5, 6, and 7]. The following notions are important in teleoperation concept.

Operator: A human operator is the person who takes the control actions needed and monitors the operated machine [4].

Teleoperator: Teleoperator is the teleoperated machine (robot). It is a machine that enables a human operator to move, to sense and to manipulate objects mechanically at a distance [4]. Most generally any tool, which extends a person's mechanical action beyond his/her reach, is a teleoperator.

Teleoperation: Teleoperation means to operate a robot using human intelligence, which requires the availability of adequate human-machine interface. Simply, the teleoperation means to operate a vehicle or a system over a distance [8].

The autonomy level of the teleoperator in the teleoperation process can be classified into the three main control classes. For visualization please see figure 1.

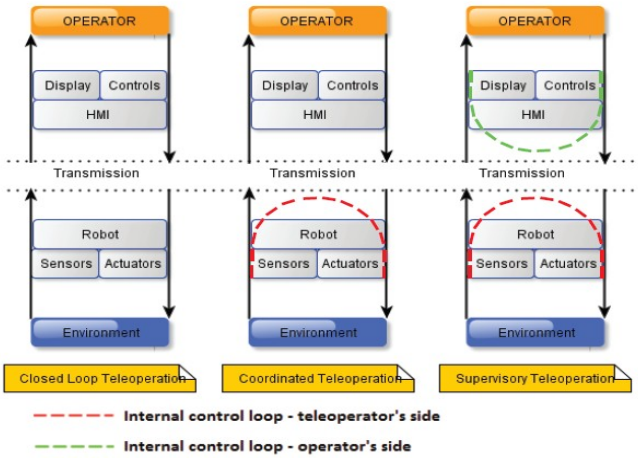


Fig. 1. The autonomy level scheme of the teleoperator in the teleoperation process derived from [9] representing the three level autonomy architecture described above.

1. Closed loop control (Direct teleoperation): The operator controls the actuators of the teleoperator by direct (analog) signals and gets real-time feedback. This is possible only when the delays in the control loop are minimal (up to 100ms). A typical example of this is a radio controlled car.

2. Coordinated teleoperation: The operator again controls the actuators, but now there is some internal control loop in

the teleoperator. However, there is no autonomy included in the remote side. The remote loops are used only to close those control loops that the operator is unable to control because of the delay. A typical example of this is a teleoperator for whom the speed control has a remote loop and, instead of controlling the throttle position, the operator gives a speed set point. Digital closed loop control systems almost always fall into this category.

3. Supervisory control: Most of the control part is to be found on the teleoperator side. The teleoperator can now perform part of the tasks more or less autonomously, while the operator mainly monitors and gives high-level commands. The term "task based teleoperation" is sometimes used here, but it is more limited than supervisory control [9].

By increasing the sensory and effectors equipment on the teleoperator side and improving the graphical user interface, so the human operator gets more and more information from the distant place, we move from standard teleoperation to telepresence, virtual presence or augmented presence operation.

A. Current Challenges in Teleoperation Systems

The abilities of teleoperation systems are being improved from year to year. Since the electro mechanical manipulator was developed, the teleoperation systems have spread to different areas of applications from military to the entertainment industry. But, what is going to happen to the teleoperation systems as we know them today? The course of the development is to decrease the human's share of control and transfer it to the teleoperator. The following picture describes the actual trend of teleoperation systems according to the iRobot Company.

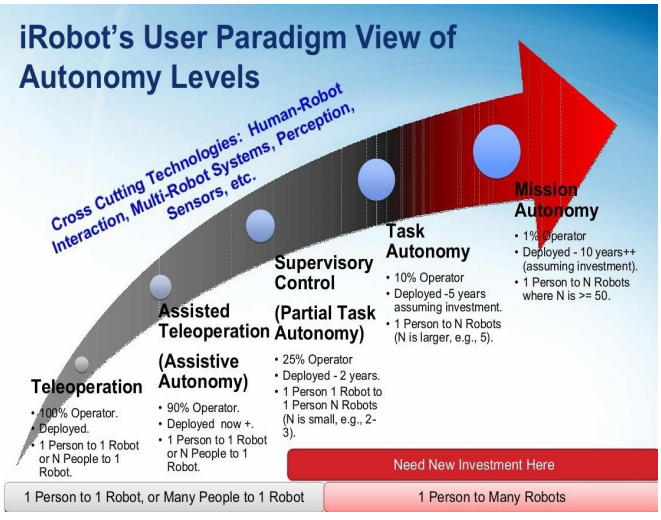


Fig. 2. iRobot's user paradigm view of autonomy levels [Robotics Summit Virtual Conference & Expo, June 2011].

Figure 2 shows the need of the transfer from manual teleoperation to task and mission autonomy of the robots. The figure 2 appeared on the Robotics Summit Virtual Conference & Expo in June 2011 and was presented by iRobot Company. According to the iRobot company the transfer from today's situation of 100%-90% human's share on the control to 1% human's share may happen in 10 years, assuming the actual level of investment in this research field. In that case, one operator would be able to control 50 and more robots in a real time, while today we do need *n* operators to control one robot. This will lead to significant lowering of expenses, errors during the task solution and human fatigue. From the other perspective of view one can disagree with the iRobot's paradigm view of autonomy levels. The main reason is that there is no common specification or standard for autonomy levels measurement. We can amend the previous picture and describe the current state in the autonomy level domain as the following picture number 6 indicates. The horizontal line of the graph (x-axis) represents the autonomy level of the robotic system and the vertical line of the graph (y-axis) represents the actual ranking of the corresponding autonomy level of the robotic system. The main problem of such autonomy level measurement is the difficulty of autonomy level determination and its ranking. While the determination of the autonomy level can be done according to the iRobot's user paradigm view of autonomy levels (picture 2), where the human's operator share on control process and the number of operators needed to control one robot are considered, the ranking of each autonomy level is dependent on different attributes. The ranking of autonomy levels might be dependent on environment complexity in which the robot can be classified into corresponding autonomy level or tasks/mission complexity as well.

For example, can an automatic lawn mower, classified as mission autonomous system, fulfill its mission if the environment is changing in time, or if the garden structure is too complex, including flowers, animals, isolated grass areas, etc...? The most common scenario is that a human intervention is needed. Thus, the operator's share on control is increasing again and autonomy level might be degraded. There one can see the connection between the autonomy levels of robotic systems and the ranking of each autonomy levels. This is the reason why more sophisticated autonomy level measurement is needed.

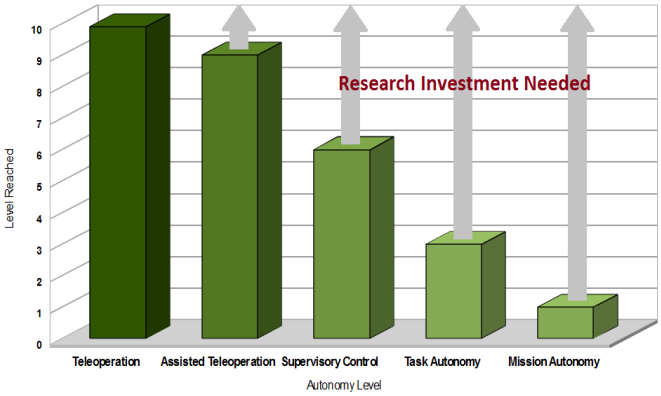


Fig. 3. Subjective description of current state of present robot autonomy levels and the reached ranking of individual autonomy levels.

It is proved that standardization of processes and measurement in general boosts up the development of technology. This is why it is so important to have a common measurement of autonomy levels in robot teleoperation systems. Autonomous robotic systems of the future should be able individually or collaboratively solve given tasks with no or only a little intervention of human to control process.

The following chapter gives a basic introduction to learning through demonstration research, which may be an important part of the transfer process from current teleoperation systems to mission autonomy systems described in this chapter.

III. Learning through teleoperation

Learning through teleoperation simply means showing a robot how to accomplish an activity in certain specific instances using a teleoperation technique. The robot generalizes the actions and is then able to perform the tasks successfully under different circumstances. However, the ability to interact with objects in a general way is still under the research process. The process of learning a mapping between world state and actions is very important in many robotics applications. This mapping enables a robot to select an action based upon its current world state and is also called a policy [10]. The development of policies by hand is often very challenging and as a result, machine learning techniques have been applied to policy development. Within learning from demonstration, a policy is learned from demonstrations or examples provided by a teacher. Learning through teleoperation is a subclass of learning from demonstration (LFD) as this can be done using the following ways of demonstration: Shadowing, sensors on teacher, external observation and teleoperation.

During teleoperation, a robot is operated by the teacher while recording from its own sensors. Since the robot directly records the states or actions experienced during the execution, the record mapping is direct [10]. Teleoperation provides the most direct method for information transfer within demonstration learning. However, teleoperation requires that operating the robot has to be manageable, and as a result, not all systems are suitable for this technique, thus some kind of smart teleoperation system is needed.

IV. Smart teleoperation system proposal

We propose a teleoperation system framework capable of operating humanoid robots, wheeled and flying vehicles and with a capability of learning from teleoperation and automation of the previously performed tasks, collaborative work of various robots during the process of task or mission solving. The following picture describes the overall architecture of the Smart Teleoperation System (STS).

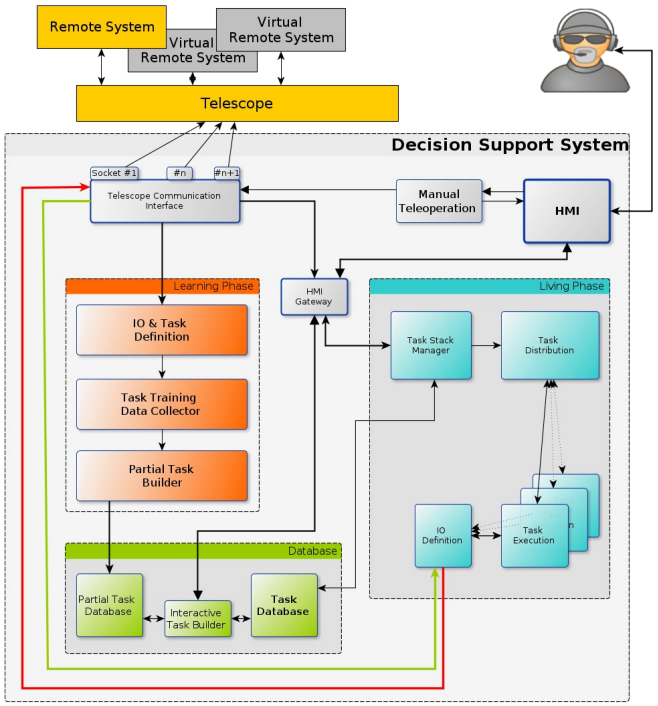


Fig. 4. The basic smart teleoperation system (STS) architecture describing the information flow between the STS, operator and teleoperator.

The main architecture consists of four parts: Human operator, teleoperators, virtual assistant and a communication interface called Telescope.

The proposed system is capable of serving one or more human or other non-human operators (other virtual assistants). Virtual Assistant can be split into another five main function parts. The standard teleoperation service provides human machine interface for the operator and provides commands transmission via Telescope to the teleoperator. In addition, there are services for learning from teleoperation, optimizing the obtained knowledge, storing it into the database and reusing the collected knowledge for assisted teleoperation.

On the right side of the picture 4, there is a main block that provides assistance to the operator during the teleoperation process. Basically the system can run in three modes: Manual teleoperation, learning mode and assisted teleoperation mode. During the manual teleoperation mode, there are no or only the little interventions to the operation of the robots and all the work has to be done by a human operator. Learning mode provides an ability of the system to learn through the teleoperation process. Assisted teleoperation mode on the other hand is working with the generalized knowledge created during the learning process and is behaving like a virtual operator. This is just an overall schematic proposal of smart teleoperation system.

A. STS Framework

Smart Teleoperation Framework is a software package built according to the proposed STS scheme. The robotic platforms for the current research so far are Aldebaran Nao, the humanoid robot and AR-Drone 2.0, the flying vehicle. The GUI of the STS framework is shown in the following picture. Robot NAO can be manually controlled using a PS3 controller. Operator simply selects the joints that need to be controlled. Then using the analog sticks of the controller, the robot is operated. In addition, the framework allows the operator to see what robot sees using the robots embedded cameras, apply any filter for the image processing or utilize other source of information (overhead cameras for example).

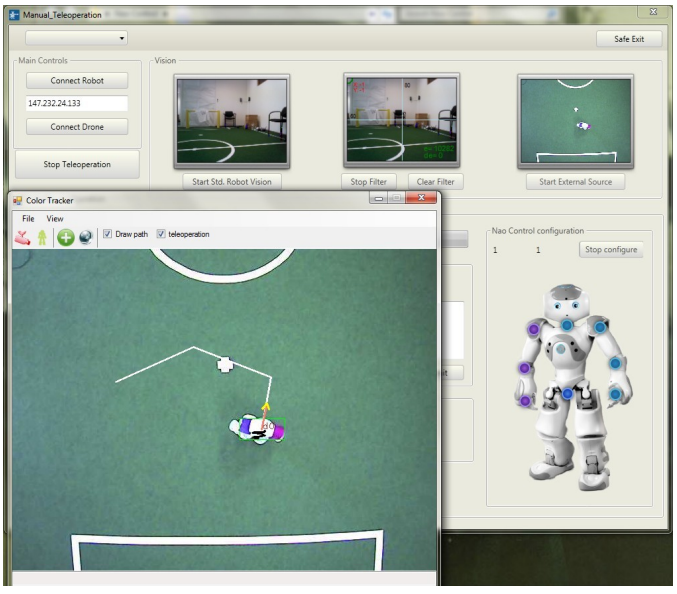


Fig. 5. The example of graphical user interface of the Smart Teleoperation System.

Operator can trigger any preprogrammed robot behaviors by pressing some combo buttons on PS3 controller and interact with the environment by using text to speech robot interface and microphones. STS framework is built as a plug-in framework so many filters for image processing can be applied, different sources of video stream can be used, or any obstacle avoidance system can be utilized with no or only little effort used.

In addition, STS framework is able to control a robot using information from overhead camera view. This can be achieved utilizing the flying vehicle or camera in laboratory conditions. Operator can control the robot like in the real-time strategy PC games. The operator simply marks the waypoints and robot then autonomously navigates to the end point. If there is any obstacle, the automatic fuzzy obstacle avoidance system (plugin for obstacle avoidance) is triggered.

B. Human to Robot Skill Transfer

The very first version of implementation of the architecture described in chapter 4 was tested by conducting some simple experiments using the STS framework and task learning plugin. The first task to be learned by the system was a ball tracker. Operator describes the properties of the ball and properties of the image processing plugin. It calculates the error of the ball and the change of the error position from the middle of the image. The input variables for the learning system were defined as the horizontal and vertical position of the ball related to the center image position. Output variables are the operator's actions performed using the controller (Nao head pitch and yaw). Input and output variables were then fuzzified. Membership function of the input and output variables were created manually using STS framework GUI. Learning process through teleoperation is triggered manually by the operator.

The STS simply monitors the operator's actions and ball position, error and change of the error in time. The rule is extracted from the every action that leads to the error lowering. For example: The ball in time *t* is *"low left and low down"*, the *error* in *t+1* was *lowered* and the operator's action was *"output"* (low right, low up) then create this rule. Fuzzy rules were represented as a fuzzy cognitive map during this experiments. After a few moments of teleoperation the system is able to track the ball automatically without the operator's intervention using the extracted rules. However for more smooth action, the further optimization is required using the training data collected during the teleoperation process and utilizing optimization algorithms.

For this experiment, the genetic algorithm was used. The full process of fuzzy cognitive maps adaptation using genetic algorithms is described in [11, 12]. The same experiment was conducted for teaching a robot to track a ball and navigate to it. The ball tracker from the previous experiment was used, and the operator focused only on the robot navigation using PS3 controller to control the walk of NAO robot. The rules for the "walk to ball" were extracted by the same way as in the previous experiment including the head angle source of information in addition. The significant lowering of the amount of operator's interventions was achieved during the second experiment, as the robot was able to track the ball automatically (learned from previous experiment) and the operator could focus on walking operation only.

V. CONCLUSION

This work is aimed to explore some possibilities for improvement of the human to robot skill transfer via teleoperation. We showed that increasing the autonomy level of the robots is very important task. In some cases, it is technologically not possible to manually operate the robotic system. If there is no or very weak connectivity the mission and sometimes the teleoperator itself is lost. This is why we focus on learning through teleoperation and increasing the task and mission autonomy of robots.

Learning from demonstration (LFD) is one of the trends in current research in this field. This work also provides an introduction to LFD methods used in human to robot skill transfer and a proposal of a novel approach to the learning through teleoperation.

In the teleoperation domain, there are a lot of technological and theoretical problems which remain to be solved. The experimental results show that the proposed architecture of the system capable to learn through the teleoperation works well. We were able to extract the rules based on the operator's actions during the teleoperation process. As the proposed system is based on the plug-in architecture, there is a possibility of extending the functionality. After teaching the virtual assistant the tasks individually, operator can connect these tasks in a sequence or in parallel run to build a mission.

The future work will be focused on further experiment work, increasing the functionality of the system, improving the graphical user interface so it is more intuitive and testing more optimization algorithms to improve the overall performance of the STS.

REFERENCES

[1] T. B. Sheridan, Space Teleoperation through Time Delay: Review and Prognosis, IEEE Transactions on Robotics and Automation, Vol. 9, No. 5, Oct. 1993

[2] M. Zhu and S. E. Salcudean, Achieving Transparency for Teleoperator Systems under Position and Rate Control, IEEE/RSJ International Conference on Intelligent Robots and System, Pittsburgh, PA August 5- 9, 1995.

[3] T. B. Sheridan, Teleoperation, Telerobotics and Telepresence: A Progress Report, Control Engineering Practice, Vol. 3, No. 2, pp. 204- 214, 1995.

[4] P. Batsomboon, and S. Tosunoglu, A Review of Teleoperation and Telesensation System, 1996 Florida Conference on Recent Advanced in Robotics, Florida Atlantic University, Florida, April 11-12, 1996.

[5] P. Batsomboon, S. Tosunoglu, and D. W. Repperger, A Survey of Telesensation and Teleopration Technology with Virtual Reality and Force Reflection Capabilities, International Journal of Modeling and Simulation, Vol. 20, No. 1, pp. 79-88, 2000.

[6] B. J. Nelson, Assimilating Disparate Sensory Feedback Within Virtual Environment for Telerobotic Systems, Robotics and Autonomous Systems 36, pp. 1-10, 2001.

[7] E. Mebarak, and S. Tosunoglu, On the Development of an Automated Design Interface for the Optimal Design of Robotic Systems, Proc. 5Th World Automation Congress, WAC 2002, and the ISORA 2002 9th Int. Symp. Robotics & Applications, Orlando, Florida, June 9-13, 2002.

[8] T. Fong and C. Thorpe. Vehicle teleoperation interfaces, July 2001.

[9] T. B. Sheridan, Telerobotics, automation, and human supervisory control. MIT Press, (1992.

[10] B. D. Argall, et al., A survey of robot learning from demonstration, Robotics and Autonomous Systems, 2009

[11] M. Paľa, L. Miženko, Mobile Robot Navigation using Adaptive Fuzzy Cognitive Maps, SCYR 2012 : proceedings from conference : 12th Scientific Conference of Young Researchers : May 15th, 2012, Herľany, Slovakia. - Košice : TU, 2012, ISBN 978-80-553-0943-9.

[12] M. Paľa, D. Lorenčík, P. Sinčák, Towards the Robotic Teleoperation Systems in Education, ICETA 2012: 10. ročník medzinárodnej konferencie o eLearningových technológiách a ich aplikáciách, 8. - 9. November 2012, Stará Lesná, Vysoké Tatry, pages 241-246., ISBN 978-1-4673-5122-5.

Software life cycle of embedded systems

[1]*Juraj VÍZI (5th year),* [2]*Veronika SZABÓOVÁ (1st year)*
Supervisor: [3]*Zdeněk HAVLICE*

[1,2,3]Dept. of Computers and Informatics, FEEI TU of Košice, Slovak Republic

[1]juraj.vizi@gmail.com, [2]veronika.szaboova@tuke.sk, [3]zdenek.havlice@tuke.sk

***Abstract*—This paper presents the architecture of the software development processes utilized in modeling of embedded systems. It focuses on modeling and analysis of the critical requirements of embedded systems in general by using the UML, MOF and XMI notation. Embedded devices are represented by tablets and smartphones.**

***Keywords*—Embedded systems, architecture, development, UML, XMI, MOF.**

I. Introduction

Embedded system is meant as applied computer system usually designed to perform a dedicated function. Embedded devices can be found on different markets, e.g. automotive, consumer electronics, medical devices, office automation etc. Few years ago they were more limited in hardware and software functionality than a personal computer (PC). These days the performance of embedded devices is comparable to PCs. The memory and processor capacitance has grown, which allows looking at these systems as on a computer itself. It means we are able to apply processes from software engineering also for these systems.

To pick up a domain for this article our view in the next text is established to mobile devices (tablets, smartphones) and medical domain. The goal is to present how the current mobile devices can be used in some medical solutions, e.g. end-user applications for patients and doctors using the knowledge of different software processes and architecture of embedded systems. The scope is also to analyze the actual representation of critical requirements of designed software system.

II. Architecture Of Embedded Systems

Generally the architecture of embed systems can be represented by a 3-tier layer model containing from following parts:

- Hardware layer
- System software layer
- Application software layer

Hardware layer contains all the physical components. By hardware in this article we will mean tablets or smartphones. Software layer is usually divided on two parts. System software can be device drivers, operating system or some kind of middleware for specific purpose, e.g. communication with database. Application software is every other software for end users, e.g. games, office or entertainment applications.

The architecture of an embedded system is an abstract representation of the embedded device. Abstract means that detailed implementation information is replaced by different objects, called elements [1]. They are representations of hardware or software with the main scope pointing to behavioral and structural information.

From point of embedded systems architecture and developing of applications it is important to have some knowledge about both software layers - system and application. The most important parts of system software layer for mobile devices are following:

- Operating system
- Middleware
- Device driver

Embedded device's operating system acts as a set of libraries, which serves as a mediator between applications from application software layer and hardware. This should make applications less dependent on hardware. The second important role of operating systems is to manage the hardware and software resources. The three most used operating systems in mobile devices are following:

- OS Android from Google
- iOS from Applet
- Windows Mobile from Microsoft

Middleware is another part of system software, except of operating system, which serves as a mediator between application software and the kernel or device driver. It can also serves as a mediator between different applications.

Top of the embedded system model is represented by application software layer. The functionality of an application is represented by this level.

III. Processes In Embedded Systems Software Life Cycle

Generally the purpose of processes in software life cycle is to define mechanisms for implementing and delivering a software product to a customer. As it is done in many projects, these processes are often adapted to specific needs of a project. This is called tailoring. The necessity of tailoring is given by different size of project teams or problem domain. Actually there are a lot of different techniques and models, which are applied in software projects. Following are some well known development models, which can be used for

embedded systems development as well:

- Waterfall
- Incremental
- Spiral
- V-model

Comparing to development methods there are development techniques, which describes the way how to implement our models. Development techniques can be:

- Prototype
- Cleanroom
- Object-oriented
- Agile techniques (e.g. SCRUM, XP)
- Lean techniques

V-model and SCRUM will be taken into account. It is an example of methodology created by a development mode (V-model) and development technique (SCRUM).

IV. Related Research

First goal is to extend the architecture of embedded systems with respect to critical requirements of developed system.

Second goal of research is to adapt the development methods in software processes.

In the initial phase of project the high level requirements document is created. These requirements are then divided to more detailed requirements, which serve as input for design. In this phase UML is often used as a reference for coding itself.

A. Usage of metamodels

This section presents some ways for representing UML diagrams by metamodels. Starting point is introducing an interchange format XMI (XML Metadata Interchange), which is used for sharing models using XML (Extensible Markup Language) [2]. It is using several aspects for describing objects in XML:

- The representation of objects in terms of XML elements and attributes is the foundation
- Since objects are typically interconnected, XMI includes standard mechanisms to link objects within the same file or across files
- Object identity allows objects to be referenced from other objects in terms of IDs and UUIDs
- Validation of XMI documents using XML Schemas

The foundation technology for describing metamodels is MOF (Meta Object Facility) [3] [4]. The MOF specification is the foundation of OMG's industry-standard environment where models can be exported from one application, imported into another, transported across a network, stored in a repository and then retrieved, rendered into different formats (e.g. XMI), transformed and used to generate application code. For UML it means that these functions are not restricted. Structural models as well as behavioral models can participate in this environment.

B. MOF and Model Driven Architecture (MDA)

Based on MOF-enabled transformations, the MDA unifies every step of the development of an application or integrated suite from its start as a Platform Independent Model (PIM) of the application's business functionality and behavior, through one or more Platform Specific Models (PSMs), to generated code and a deployable application. The PIM remains stable as technology evolves, extending and thereby maximizing software ROI (Return on Investment). Portability and interoperability are built into the architecture.

MDA relies on the MOF to integrate the modeling steps that start a development or integration project with the coding that follows.

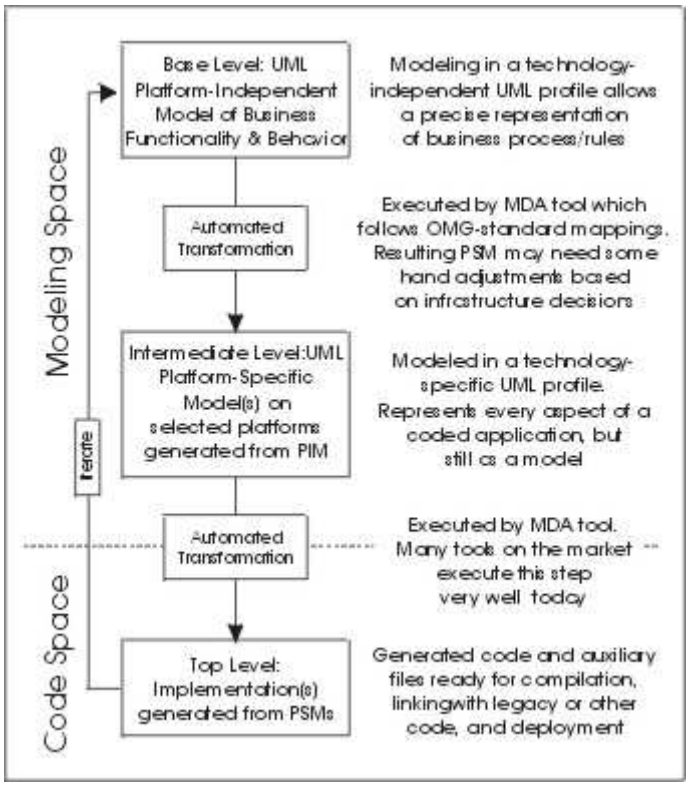


Fig. 1. MDA Specification

C. UML Infrastructure Architecture

The Infrastructure of the UML is defined by the InfrastructureLibrary, which satisfies the following design requirements [5]:

- Define a metalanguage core that can be reused to define a variety of metamodels, including UML, MOF
- Architecturally align UML, MOF and XMI so that model interchange is fully supported
- Allow customization of UML trough Profiles and creation of new languages (family of languages) based on the same metalanguage core as UML

Infrastructure is represented by two packages:

- *InfrastructureLibrary*
- *PrimitiveTypes*

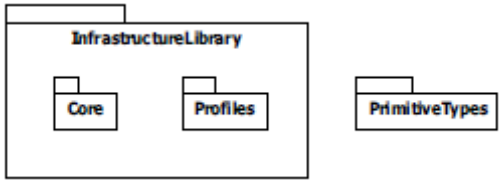


Fig. 2. The InfrastructureLibrary package

The *InfrastructureLibrary* package consists of the packages *Core* and *Profiles*. The *Core* package contains core concepts used when metamodeling and the *Profiles* package defines the mechanisms that are used to customize metamodels.

One of the major goals of the Infrastructure has been to architecturally align UML and MOF. Several approaches were defined for this. The first approach has to define the common core, which is realized as the package Core, in such a way that the model elements are shared between UML and MOF. The second approach has been to make sure that UML is defined as a model that is based on MOF used as a metamodels as illustrated on the picture.

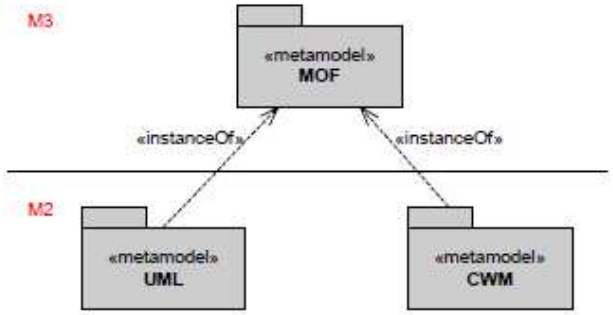


Fig. 3. UML and MOF at different metalevels

D. *XMI Model*

The XMI model is an instance of MOF for describing the XMI-specific information in an XMI document, such as the version, documentation, extensions and differences, as shown in Fig. 4:

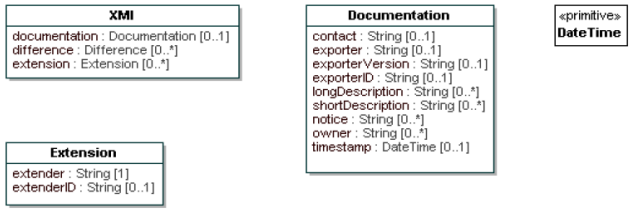


Fig. 4. The XMI Model for the XMI element, documentation and extension

The declaration of XMI:

```
<xsd:complexType name="XMI">
  <xsd:choice minOccurs="0" maxOccurs="unbounded">
    <xsd:any processContents="strict"/>
  </xsd:choice>
  <xsd:attribute ref="id"/>
  <xsd:attributeGroup ref="IdentityAttribs"/>
  <xsd:attributeGroup ref="LinkAttribs"/>
  <xsd:attribute name="type" type="xsd:QName" use="optional"
                 form="qualified"/>
</xsd:complexType>
<xsd:element name="documentation" type="Documentation"/>
<xsd:element name="difference" type="Difference"/>
<xsd:element name="extension" type="Extension"/>
<xsd:element name="XMI" type="XMI"/>
```

Fig. 5. XMI Declaration

For representing the most critical requirements UML can be used, providing diagrams for representing behavioral and also structural properties. Several applications, which belong to the same workflow, can have similar critical requirements. If these requirements are described by UML diagrams and transformed to metamodels, then could be transferable to other applications design within the same workflow. In medical domain it can be represented by applications from CT and radiotherapy. These systems contain a lot of critical requirements and are part of one big workflow.

V. Conclusion

The increased performance of embedded systems leads to improving processes in software life cycle. This allows performing many operations on embedded devices, which were in past limited because of hardware limitations. This paper mainly describes the embedded systems design. This phase is a subject of research and with help of UML provides some ways for improvements.

Acknowledgment

This work was supported by the Slovak Research and Development Agency under the contract No. APVV-0008-10 and by the Cultural and Educational Grant Agency of the Slovak Republic, Project No. 050TUKE-4/2013: “Integration of Software Quality Processes in Software Engineering Curricula for Informatics Master Study Program at Technical Universities Proposal of the Structure and Realization of Selected Software Engineering Courses”.

References

[1] T. Noergaard, “Embedded Systems Architecture – A Comprehensive Guide for Engineers and Programmers”, Elsevier, ISBN: 0-7506-7792-9, pp. 20-27, 398, 460-461

[2] OMG MOF 2 XMI Mapping Specification, http://www.omg.org/spec/XMI/2.4.1/PDF, pp. 13, 17

[3] OMG’s MetaObject Facility, http://www.omg.org/mof/

[4] OMG Meta Object Facility (MOF) Core Specification, http://www.omg.org/spec/MOF/2.4.1/PDF, pp. 11

[5] OMG Unified Modeling Language (OMG UML) Infrastructure, http://www.omg.org/spec/UML/2.4.1/Infrastructure/PDF, pp. 11-12

Solution for broadcasting video with best shot of tracked object captured by multiple mobile phones in real time (May 2013)

[1]*Dávid CYMBALÁK (2st year)*
Supervisor:[2]František JAKAB

[1,2]Department of Computers and Informatics, FEI TU of Košice, Slovak Republic

[1]david.cymbalak@cnl.sk, [2]frantisek.jakab@tuke.sk

***Abstract*—This work presents the idea of solution for broadcasting the video from multiple mobile phones with the best shot of the tracked object. Based on analysis of appropriate mobile streaming and mobile based object tracking technologies, there was outlined the concept of interconnection tracking, detection and stream control mechanism for mobile platforms. The overall solution is capable to controlling of video stream from multiple integrated cameras in mobile devices, where it is also simultaneously calculated the coordinates of tracked object by TLD algorithm.**

***Keywords*— Mobile Broadcast, Object Tracking, OpenTLD, FastCV.**

I. Introduction

In recent years, multimedia applications and streaming technology have become so popular that it generated more than half of the total data flow on Internet. Nevertheless, the streaming technology transfers the data over IP network without any special or extended support by the network. This situation is possible by the advanced development of highly efficient video compression techniques, the availability of broadband access technologies, but also through the development of adaptive streaming, that can compensate for the unpredictability of the network by sophisticated methods such as adapting to available bandwidth, caching methods or techniques of masking and error recovery [1]. This features are very important for streaming from or in to mobile platforms.

With increasing of demand for social networks with streaming capability, it is very interesting to provide new solutions in this field mostly based on interconnection of existing common technologies. The goal of this work is to introduce a concept of solution which brings social characteristic to delivery video stream from live event with some tracked object of interest. The live broadcasts could be created by people with mobile phones capturing same scene in the event, but the resulting solution would deliver only one video stream with best shot of the tracked object of interest to the video player windows on web portal.

II. Broadcasting from Mobile Devices

A. Streaming technologies related to mobile platforms

Mobile streaming in practice could be considered as streaming multimedia content over wireless networks or mobile network operators (2G, 3G or 4G). In case of proposal of solution, it will be considered that the mobile device is not the common receiver of the video stream, but it is the broadcaster of video stream. It can be a mobile phone, tablet or PDA with built-in camera or series of cameras.

The ability to realize video streaming from mobile device, or ability to connect to streaming server through a mobile device by the wireless networking technology creates several problems that need to be addressed. Significant factors affecting the streaming of multimedia content are differences in the network, which transmits the multimedia content. The significant impact on the mobile streaming could consider also factors such as the bandwidth, delay, jitter, packet loss and noise. For that reasons it is important to choose the appropriate streaming platform, codecs or protocols, which is optimized for mobile devices. Nowadays, the most used mobile streaming codec is H.264/AVC.

Advanced H.264/AVC is codec developed by the ITU-T VCEG and ISO / IEC MPEG. Codec is composed of two layers: video coding layer (VCL) and network abstraction layer (NAL). VCL represents the actual content of the video and NAL cares about the structure of the data and carries the information necessary for the transmission, which tend to be used by the transport layer and the storage media [2,3]. H.264 offers several enhancements to support streaming. A useful addition is a flexible sort of macroblocks (FMO) [4]. The most common file container for H.264 codec is MP4 file format. Scenes in the MP4 file format consists of a hierarchical multimedia objects. The first is layer of static images that form the background of the scene, a second is layer of the moving objects, and a third is layer of sound objects. Individual sub-streams between source of streaming and destination are managed by synchronization layer and

delivery layer, which is divided into sublayers DMIF and TransMux [5]. HEVC Codec H.265 or MPEG-H Part2 is a highly efficient video codec developed as a continuation of the popular H.264 codec by JCT-VC group (merger of VCEG and MPEG). Compression has been improved in high level that in comparison with the original H.264 it will suffice files half the size and half the bandwidth requirements for streaming. This type of compression will be used in for video encoding of the future standards of high-definition video up to resolution of 4K or 8K. Unlike H.264, H.265 HEVC will contain only one main section, which should ensure greater compatibility between standard and mobile devices [6].

Another side, what could be consider for optimize mobile streaming technologies, could be choosing of appropriate streaming protocol. RTMP (Real Time Messaging Protocol) is a protocol based on TCP, which maintains a permanent connection and allows communication with low latency. For flawless and efficient delivery of content, stream is split into fragments whose size is negotiated dynamically between the client and the server. HTTP streaming is the possibility of progressive delivering of data. Conventional processes send HTTP response to the HTTP request based on a fixed size header. The table of contents is therefore necessary to know its size before initialization and then it can be integrated in response to the request [7].

B. Applications for broadcasting from mobile phones

Streaming from mobile device with integrated camera can be realized by connecting from recipient to the network address of the mobile device and pulling data (PULL streaming type). Another way is to establish a connection from mobile device to the streaming server, which redistribute and send streaming data to recipient (PUSH streaming type). The first option (PULL) is suitable for unicast connection with one or small number of stream recipients, where there is no overhead for bandwidth or performance mobile devices components. The second option (PUSH) is suitable for receiving stream by multiple recipients, who are able to receive live broadcast via streaming server, which serves as a redistribution mechanism for receiving unicast stream in real time from your mobile device and simultaneously sending unicast or multicast streams to the recipients. For design of solution outlined in this work it would be consider using mobile devices with Android operating systems. There were tested applications on Android platform with streaming ability from integrated camera such as C++ RTMP Server, Broadcaster or IP Webcam.

C++ RTMPD Server supports x86, amd64, PPC, ARM and MIPS platforms. It was designed primarily for Unix systems, but it is able to run on the Android platform. Server supports RTMP, RTMP, RTMPS, RTMPT, RTMPTE with ability of broadcasting from Android or iOS devices or working with IP cameras.

Broadcaster app is designed for operating systems Android and Blackberry. It implies encoder that can encode real-time video using H.264 or H.263 (VP6) codecs and using Nellymoser for audio. The output stream can be sent to the streaming server of PUSH type as Wowza, Red5 or mentioned C++ RTMPD. For video streaming using RTMP, RTMP, RTMPS, RTMP there is also support for authentication of connecting to the streaming server. It has options for the resolution, codec and frame rate. It is able to select video source between front and rear camera of mobile device [8].

Another streaming application - IP Webcam - allows using the mobile device with Android as IP network camera, which acts as a streaming server type PULL using the HTTP protocol. The program can adjust the quality of the streamed image, size, resolution, set the username and password for security, set the streaming port or implement secure transmission via HTTPS. The application creates a server on port 8080, which is also a web server reachable from web browser. With this technology it is possible to use MJPEG software that processes the stream of mobile devices and creates a virtual webcam driver, which is represented as normal camera device in computer on recipient side [9].

III. Object Tracking Solutions for Mobile Devices

Nowadays, there is a couple of approach for tracking objects in video, but there is a lack of implementation for mobile platforms. The TLD algorithm and its implementation OpenTLD have shown as reliable and effective in the combination with FastCV library for applying computer vision methods.

A. OpenTLD C++

TLD (Tracking-Learning-Detection) algorithm is developed to detect and track objects in the video in real time [10]. One of the implementation of TLD algorithms is OpenTLD in C++ using only open libraries without additional Matlab environment [11]. The resulting application has an option to define a template for object tracking either the manual way, or the object model, or by initialization coordinates. It includes option for running with the parameter defining the video source for object tracking in real time or from video recording. It has ability to define the output file with history of coordinates of tracked object position and it also store the learned model of the observed object. Learned object model can be used in a subsequent tracking or in other instances of program with other video sources.

There is also OpenTLD application on Android platform, which is created by OpenTLD C++ implementation in cooperation with FastCV library that provides a live preview rendering with displaying the tracking layer around the detected and tracked object [12]. The application operates on the principle of defining a template for detection on first rendered frame from camera. Then it starts the detection of template model in the actual picture frame in real time, where another characteristic of tracked object is learned and extended in to template model (Fig. 1). The application is only a prototype, which is used to label tracked object by drawn rectangle in the live preview of video. An interesting

addition to this application would be to extend the opportunity to continuously gather information about the position and size of the object in the form of coordinates.

B. *FastCV library*

Computer Vision Library FastCV was developed by Qualcomm and later it has been released for free using. Its main advantages include: functions for image processing and transformation, detection, tracking, recognition, 3D reconstruction functions for use in Augmented Reality, segmentation and advanced memory management. FastCV is primarily designed for mobile devices, ARM platform, in which it achieves significantly faster and more efficiently results than competing image processing library such as OpenCV [13].

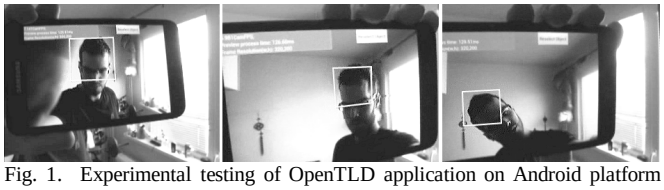


Fig. 1. Experimental testing of OpenTLD application on Android platform with mobile phone device with integrated camera, which tracks the human head based on similarity with head model template.

IV. Concept of Solution for Broadcasting Best-match Video of Tracked Object

Proposed system would consist of detection, tracking, evaluation and stream control mechanism with connection to streaming server and analyzed streaming and tracking application on android platform (Fig. 2).

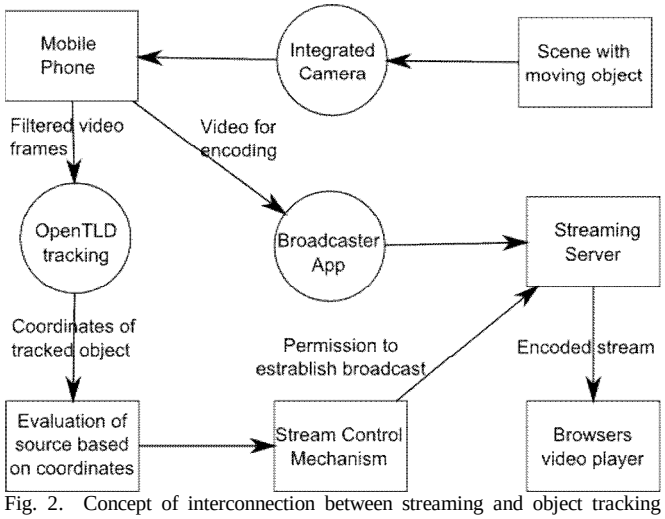


Fig. 2. Concept of interconnection between streaming and object tracking technologies for one mobile phone.

The idea of solution is the existence of a social web portal with events, or places, from where you can watch a moving object of interest in scene (for example event with performance of persons located on podium captured by people with mobile devices - Fig. 3). Each mobile client with broadcasted video would be possible to subscribe to the selected event, or it would be able to create a new event, which would be limited to a specific geographic location, and for which it would have to define an initial template model for tracking of object of interest. The application should send

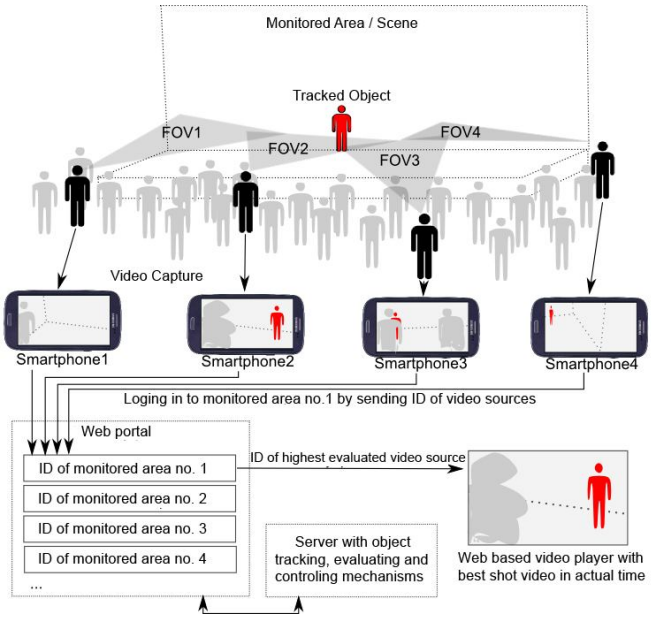


Fig. 3. Visualisation of designed solution with four connected smartphone from global view with interconnecting to social event based web portal

the ID of mobile device to the system, which would be stored in the database of streaming resources for establishment the streaming connection to recipients on web portal. The solution could use streaming application like IP webcam and broadcast video from the integrated camera in real-time. At the same time the application would be able to process the video frame in real time using the FastCV library and object tracking methods such as TLD. The solution counts with the ability to distribute the learned object template model for object tracking to other devices registered for the event. Evaluation based on actual coordinates from each streaming source would be regularly sent to the server with the mechanism of evaluation metrics and stream control mechanism, which decides, which streaming source will be redistributed to the recipients on web portal. The users of web portal would got the opportunity to watch the event through the views from mobile devices of event participants. The best shot on the tracked object of interest will be always chosen from the outgoing streams in the real time. This will display the object in the best evaluated compositions, sizes and credibility in actual time. The resulting video stream is played by the user in a web player with cooperation with streaming server with actual stream ID, which at that time represents the best video source.

V. Conclusion

In terms of deployment and implementation of the proposed system in an environment of mobile devices would be necessary to optimize each from streaming, detection and tracking mechanisms in order for their reliable operation with sufficient performance of state of arts smart mobile phones. However, it could assume in the near future, the mobile devices will reach new levels of performance that will not have a problem with the deployment of streaming server directly to the device at the same time with detection and tracking technology. In addition it could bring the efficiency

in tracking and evaluating source rating from object coordinates, when it will be possible to set shorter intervals in selecting of relevant video frames for detection.

ACKNOWLEDGMENT

Research described in the paper was done by cooperation with Computer Networks Laboratory at DCI FEEI TU in Košice. This work is the result of the Project implementation: Competency Centre for Knowledge technologies applied in Innovation of Production Systems in Industry and Services, ITMS: 26220220155, supported by the Research & Development Operational Programme funded by the ERDF.

REFERENCES

[1] AKHSHABI S. - BEGEN A. - DOVROLIS C.: And Experimental Evaluation of Rate-Adaptation Algorithms in Adaptive Streaming over HTTP, MMSys'11, San jose, USA, 2011
[2] DHONSALE K.V. : Overview, implementation and comparison of Audio Video Standard (AVS) China and H.264/MPEG -4 part 10 or Advanced Video Coding Standard, 2012
[3] WIEGAND T. and etc..: Overview of the H.264 / AVC Video Coding Standard, IEE Transactions on Circuits and Systems for Video Technology, 2003
[4] DHONDTA Y. and etc.: Flexible macroblock ordering as a content adaptation tool in H.264/AVC,Multimedia Systems and Applications VIII, Proc. of SPIE Vol. 6015, 2005
[5] ISO/IEC JTC1/SC29/WG11 CODING OF MOVING PICTURES AND AUDIO: Overview of the MPEG-4 Standard, ISO N4668, 2002
[6] SULLIVAN G. and etc: Overview of the High Efficiency Video Coding (HEVC) Standard, IEEE trans. On Circuits and Systems for Video Technology, 2012
[7] DAOUST F. and etc: Towards Video on the Web with HTML5, NEM Summit, Barcelona, 2010
[8] Broadcaster APP [online], available on internet: https://play.google.com/store/apps/details?id=air.Broadcaster&hl=en
[9] IP Webcam Documentation [online], available on internet: http://ip-webcam.appspot.com/
[10] KALAL Z. - MATAS J. - MIKOLAJCZYK K.: P-N Learning: Bootstrapping Binary Calssifiers by Structural Constraints, 23rd IEEE Conference on Computer Vision and Pattern Recognition, CVPR, San Francisco, USA, 2010
[11] NEBEHAY G.: RObust Object Tracking Based on Tracking-Learning-Detection, Diplomarbeit, Technische Universität Wien, 2012
[12] GRAVDAL E.: Augumented Reality and Object Tracking for Mobile Devices, NTNU Trondheim, 2012
[13] QUALCOMM INC.: FastCV Library 1.1.1 [online], available on internet: https://developer.qualcomm.com/docs/fastcv/api/index.html

Spiking Neural Networks: Introduction and State-of-the-Art

[1]*Radoslav BIELEK (1st year)*
Supervisor: [2]Ján JADLOVSKÝ

[1,2]Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

[1]radoslav.bielek@tuke.sk, [2]jan.jadlosvky@tuke.sk

***Abstract*—We want to introduce the field of the Spiking Neural Networks and basic knowledge necessary to know for its understating on the main basis. Paper is divided into sections dealing with biology of neuron, history of artificial neural networks, neuron models, learning algorithms and so far achievements obtained with the future vision.**

***Keywords*—Spiking neural networks, neuron models, self-organizing systems.**

I. Introduction

To get an insight into the field of Spiking Neural Networks (SNNs) it is important to obtain information from more fields. Especially the neuroscience to have an outline of the electrical properties of the real neuron and to learn about classic neural networks (NNs) as the many principles remained unchained.

II. Biological Background

Single neuron cell is created of the neuron body – soma, the inputs to the neuron called dendrites and the outputs called axons (Fig. 1).

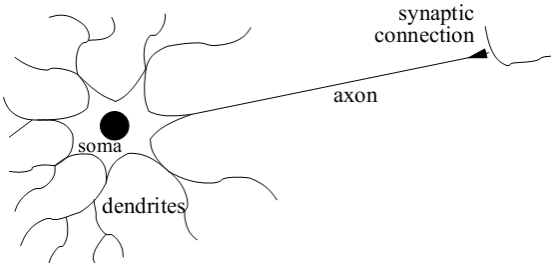


Fig. 1. Scheme of the biological neuron. Depicted from [1].

The main interest in the field of Spiking Neural Networks is the spreading of the electrical signal through the whole neuron and its propagation in the networks of multiple neurons together with the interpretation of these signals. The connections between neurons are called synapses and are created by the end of the axon and beginning of the dendrite.

As it is mentioned above the dendrites bring the electric signals from the previous neurons (called pre-synaptic neurons). These signals are integrated and are of two types: inhibitory postsynaptic potential (IPSP) or excitatory postsynaptic potential (EPSP) according to the effect they have. If the final sum of them reaches or exceeds the threshold value, the neuron itself generates action potential of short lasting (1 ms) which are called spikes [19].

Spikes are generated as a result of equalizing the conductance between the interior and exterior of the neuron. This is happened by the closing and opening the ion channels placed in the impermeable lipid bilayer [11].

Spikes are the main means of communication among neurons. The main debate of neuroscientists consists of whether is the information in the brain coded by the frequency or the rate of the spikes, but the most probable is the combination of these two views on the problem. Hence, the most important information is the spike generation itself.

After the spike generation the threshold value is raised to the not reachable values and neuron comes to the so-called refractory period (10 ms) in which there cannot be any new action potential generated.

The generated action potential is spread down the axons to the followed postsynaptic neurons. The complexity of the human brain is made up due to the enormous number of inter-neural connections. One neuron can receive signals from up to 10 000 synapses and pass them to the other thousands of neurons [19].

The multiple types of the spike behavior (Fig. 2) also add to the information coding complexity as they can vary stochastically.

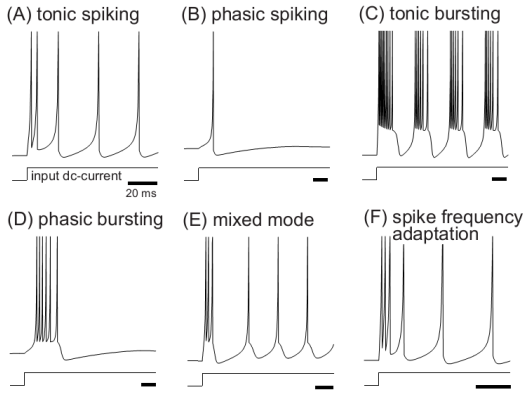


Fig. 2. Examples of the neuron behaviors. In fact there are up to 20 various ways in which neurons can generate spikes. Depicted from [1].

The number of the brain neurons in the animal kingdom starts on 10 000 (medical leech), continues with 960 000 in the case of honey bees and ends with the massive brain of elephant which is built from 23 000 000 000 neurons [18]. Humans as the beings with most developed brain on the planet

have brain that consists of 10^{11} – 10^{14} neurons with approximately 10^4 synaptic connections on each neuron [2].

III. History of Neural Networks

The SNNs are also referred to as 3rd generation of the NNs. The main difference with the previous generations is in the emphasis of the biological resemblance with the real neurons. While computing with spikes many mechanisms of the SNN work had to be changed or even reimagined. Classic NNs were also inspired by the nature, but only in the very loose way.

The first neuron was called the binary neuron and was proposed by the McCulloch and Pitts [9] and its inputs and outputs were strictly binary. It was showed that any of the Boolean function can be implemented with these computational units. The connections with the biological neurons are in this case very loose where the synaptic weights are implemented only as the connections without expression of the synapse strength and the outputs of the network were not the spikes, but binary values. These NNs are often called as the NNs of 1st generation.

The 2nd generation of the neurons used not only binary values, but the range of real values (mostly between the 0 and 1). This modification enlarged the usage possibilities and strengthened up the similarity with the life neuron (as the synaptic connections could be evaluated by the real number) [2]. The simplest neural network from this generation is called perceptron and was invented by Frank Rosenblatt [10].

IV. Neuron Models

The choice of the neuron model to work with represents an important task. The main parameters are the biological resemblance to the real neuron (if the aim is to simulate) and the computational demands. As there are many of the neuron models with various complexity this paper describes the most used ones in general as its aim is to give reader a short insight to the SNN and corresponding equations would exceed its range. More about neuron models can be found in [1], [11], [12], [13] and [17].

A. Models using differentials equations

1) Hodgkin-Huxley Model

The one of the most precise models of the neuron's electric properties is known over half of the century and it is very well presented in the [1] and [11].

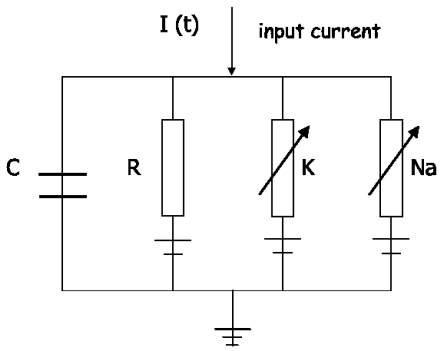


Fig. 3. Hodgkin-Huxley model in the form of electrical circuit. Depicted from [1].

It is called Hodgkin-Huxley model after its invertors and the main equation describes the electrical behavior of the neuron by the simulating of the ion channels' permeability. It is considered as one of the most accurate model and many simpler ones are derived from it. The circuit used for its simulation is shown in the Fig. 3. Simulation of this model on the computer is represented by the solution of differential equations used for describing this circuit.

Together with the accuracy of the model rises also the computational power needed to compute with it. If we imagine neural network with thousands of the spiking neurons and even more connections among them it is very suitable to work with the simplified model.

2) Integrate-and-fire Models

The basic Integrate-and-Fire (I-and-F) model (Fig. 4) was proposed long before the mechanism of action potential was discovered by Lapicque in 1907 [11]. In fact it is derived from the equations of Hodgkin-Huxley model [1]. Even though the biological relevance of the model is quite poor (Fig. 5) the I-and-F model is very often used because of its low computational power demands.

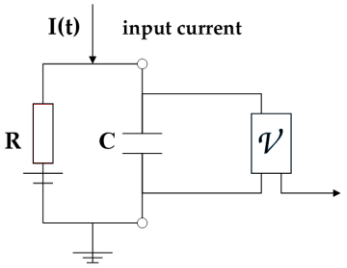


Fig. 4. Electrical circuit of the Integrate-and-Fire mode. Depicted from [1].

3) Comparison

In the process of creation SNN creates the choice of the neuron model important step. Two main parameters are the computational power and the biological resemblance with the real neuron. Comparison of the most used neurons (which use differential equations) from these views is shown in the Fig. 5.

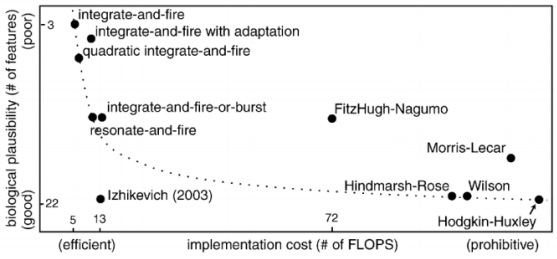


Fig. 5. Neuron models comparison from the points of the biological resemblance with real one and the computational power needed for simulation (FLOPS - number of floating points necessary for simulation during 1 ms). Depicted from [12].

B. Other models

Both models described above were created with the differential equation. The Spike Response Model (SRM) simulates the neuron behavior with the integration of the past events and the use of simple sums makes this model easier to implement in comparison with ones mentioned above. More information can be found in [13].

Special mention belongs to the CoDi model of the neuron (from collect and distribute). It is cellular automata where cells represent empty spaces, neuron bodies, axons or dendrites. The functioning of this neural model is based on a few simple rules and it is divided to two life phases: growth and signaling. As it can be deduce the individual neurons are very simple and they are not able to provide other behavior than generating of single spikes. On the other hand the simplification brings the very high performance and to adapt and learn this SNN evolutionary algorithm can be used in which the competing individuals are the isolated SNNs.

The authors of [3] and [4] are trying to construct specialized CoDi hardware (evolvable) or to implement it on the still more popular FPGA (Field-Programmable Gate Array) to fully utilize the parallel functioning of the SNN. By this way the mentioned authors believe to start new scientific field called "brain-modeling". Presented achievements of the CoDi were simple timer and sine wave generator.

Project of CoDi shows that even the very simplified models of the neurons can bring interesting results and they can compensate by the speed of computation what they are not able to offer in the field of simulation.

V. Learning Algorithms

There are several tendencies to adapt learning from previous NN generations (mainly the backpropagation, see [1]), but the basic inspiration of the SNN learning process also came from the nature. Main principle of the neuron organization in the brain is described by the Hebb's rule:

When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased [14].

The SNNs are not the first neural networks which are organized this way. The main use of Hebb's rule can be found in the Hoppfield's network [15].

That concludes to the learning without teacher when the whole neural network is self-organized according to the neurons' positioning and the signal spreading.

The process of described metabolic change is known as synaptic plasticity and according to the lasting of this change and the effect (strengthening or weakening the inter-neural connections) we can differ 4 types of them: Long Term Potentiation (LPT), Long Term Depression (LTD), Short Term Potentiation (STP) and Short Term Depression (STD). The long terms have the time effects of several hours and short ones last from few seconds to several minutes [1].

Big interest is dedicated to the Spike-Timing Dependent Plasticity (STDP) or also called temporal Hebbian rule and it shows that the signal can be spread not only from axon to dendrite, but the other way too so it creates some kind of backpropagation.

STDP shows that not only excitation of neurons is important but the time of the generated spikes also. The most significant reinforcement of the synaptic weight is present when the presynaptic neuron generates spike short time before the postsynaptic one [1].

VI. Achievements

A. Spiking PID controller

Considering the PID controller as the basic element used for regulation, the possibilities of SNNs usage in cybernetics are limited as long as they could not be able to equalize the performance of PID regulator.

The authors of [21] tried to approximate and substitute each of the PID elements by one spiking neuron (using Izhikevich's model) creating simple network consisting of three neurons. Input current to neurons was mapped to the value of error in the control loop.

Results obtained showed that the spiking neurons can approximate the functions of proportional, integral and derivative terms of PID controller, but further research has to be done especially focusing on the more complex networks.

One advantage of spiking PID controller can be relatively easy hardware implementation even using the most complex neuron model known.

B. Robot Control

Interesting experiments in the robot control with the use of computer vision were carried out by the author of [8]. He has built fully autonomous robot with two cameras, two engines and two SNNs were used for processing the signals from sensors and for motion control each one controlling the half of the robot. The outputs from these SNNs were processed in the selector and overall motion control was conducted by the more active SNN. Learning of the networks was acquired by the genetic algorithm with individuals representing the topology and connections in the network. This way of the control was inspired by the vision of insects and their separate fields of vision for each eye.

Similar experiment was held also by the [6]. Results from these works show that the genetic algorithms are very perspective way in the SNN learning.

Using the STDP principles to train the SNN in the purposes of control brings also the respectful results. In [7] the robotic arm with 4 degrees of freedom was controlled by the SNN made of Izhikevich's neuron model.

C. Human Machine Interface

Scientific field which asks for the use of the SNNs is the human machine interface creation especially the one with interconnection between machine and brain. The aim is to measure and interpret the activity of human neurons and use it to control i.e. prosthetic arms in order to substitute missing limb. This kind of research is described i.e. in [5].

D. Other Usage

Many of the researchers are trying to substitute the classic NNs in order to show that SNNs can equalize them from the point of computational power. This can be seen in the fields of computer vision and image recognition [16], continuous function approximations [17] or even for storing and retrieving information [20].

VII. Conclusion

The importance of the SNN lies in the better understanding of the human brain and with better understanding we should

be capable of not only simulate or imitate the human thinking, but to create the real minds capable of reasoning or even ones blessed with the creativity.

The most futuristic vision is that one day we would be able to interpret the spikes of the real human brain, but this would not be possible in decades, as we are not able to observe all of the neural activity in the human brain even with the most developed technology as i.e. fMRI.

The next step in the field of the neural networks could be the including of the neural chemical properties to bring the possibility i.e. communicate in various time levels or to simulate effect of the chemical substances on the human brain, but the most important task for the SNNs remains to better understand ourselves by understanding our minds.

VIII. Acknowledgement

This work has been supported by the Scientific Grant Agency of Slovak Republic under project Vega No.1/0286/11 Dynamic Hybrid Architectures of the Multiagent Network Control Systems.

References

[1] H. PAUGMAN-MOISY, S. N. BOHTE, "Computing with Spiking Neuron Networks" in *Handbook of Natural Computing*, Heidelberg, Springer Verlag, Germany 2011.

[2] P. SINČÁK, G. ANDREJKOVÁ, *Neurónové siete – inžiniersky prístup,* Košice, ELFA Press, 1996.

[3] F. GERS, H. de GARIS, M. KORKIN, *CoDi-1Bit: A Simplified Cellular Automata Based Neuron Model*, 1997.

[4] F. GERS, H. de GARIS, M. KORKIN, *"CoDi-1Bit" A Cellular Automata Based Neural Net Model Simple Enough To Be Implemented In Evolvable Hardware*, 1997.

[5] J. DETHIER, P. NUYUJUKIAN, C. ELIASMITH, et al., "A Brain-Machine Interface Operating with a Real-Time Spiking Neural Network Control Algorithm", in *Neural Engineering, 2011 5th International IEEE/EMBS.* 2011.

[6] H. HAGRAS, A. POUNDS-CORNISH, M. COLLEY, V. CALLAGHAN, G. CLARKE, "Evolving Spiking Neural Network Controller for Autonomous Robots" in *Proceedings of ICRA, p. 4620-4626, 2004.*

[7] A. BOUGANIS, M. SHANAHAN, "Training a Spiking Neural Network to Control a 4-DoF Robotic Arm Based on Spike Timing-Dependent Plasticity" in *WXXI 2010 IEEE World Congress on Computation Intelligence.* Barcelona (Spain), 2010.

[8] P. TRHAN, "The Application of Spiking Neural Networks", *Autonomous Robot Control*. Computer and Informatics, 2010, vol. 29, p. 823-847.

[9] W. S. McCULLOCH, W. PITTS, "A Logical Calculus of the Ideas Immanent in Nervous Activity" in *Bulletin of Mathematical Biophysics 5,* p.115-133, 1943.

[10] F. ROSENBLATT, "The Perceptron: A Perceiving and Recognizing Automaton", *Psychological Review*, vol. 65, 1958.

[11] P. DAYAN, L. F. ABBOTT, *Theoretical Neuroscience*. MIT Press, 2005. ISBN 0262541858.

[12] Z. FOUNTAS, "Spiking Neural Networks for Human-like Avatar Control in a Simulated Environment", Ph.D. dissertation, Dep. of Computing, Imperial College London, London, 2011.

[13] W. GERSTNER, W. M. KISTLER *Spiking Neuron Models. Single Neurons, Population, Plasticity*, Cambridge, Cambridge University Press, 2002. ISBN 05211890799.

[14] D.O. HEBB, *The Organization of Behavior*. Wiley, New York, 1949.

[15] D. ĎURAČKOVÁ, *Od biologického neurónu k integrovanému obvodu*, Bratislava, Slovak University of Technology, 2004. ISBN 80-227-2040-2.

[16] A. GUPTA, L. N. LONG, "Character Recognition using Spiking Neural Networks" in *Proceedings of International Join Conference on Neural Networks*, Orlando (Florida, USA). 2007.

[17] W. MAASS, W. "Computing with Spiking Neurons" in *The Handbook of Brain theory and Neural Networks*, MIT Press, Cambridge, 2003.

[18] *List of animals by number of neurons* [online]. Available: http://en.wikipedia.org/wiki/List_of_animals_by_number_of_neurons

[19] J. MYSLIVEČEK et al. *Základy neurověd*. Prague, TRITON, 2009, ISBN 978-80-7387-088-1.

[20] M. ABELES, *Corticonics: Neural Circuits of the Cerebral Cortex*, Cambridge, England, Cambridge University Press, 1991. ISBN 0521376173.

[21] A. WEBB, S. DAVIES, D. LESTER, "Spiking Neural PID Controllers" in *Proceedings of ICONIP*, 2011, pp.259-267.

Switching of video stream source from several cameras based on position of tracked object in actual video frame (May 2013)

[1]*Dávid CYMBALÁK (2st year)*
Supervisor:[2]František JAKAB

[1,2]Department of Computers and Informatics, FEI TU of Košice, Slovak Republic

[1]david.cymbalak@cnl.sk, [2]frantisek.jakab@tuke.sk

***Abstract*—Paper deals with the analysis of streaming and object tracking technologies with ability to work with several cameras with switchable output based on tracked object position. Goal of this work is to design a concept of stream control mechanism with tracking system with ability to detect object in real time across several cameras. Proposed system has ability to stream video from one source with the best view on tracked object in real time. As evaluation parameter for stream control mechanism it was outlined a metric for evaluate each node of multi-cameras system, which is related to current tracked object position, size or reliability of object detection. There was also outlined the concepts of interconnection of solution based on PULL or PUSH streaming servers.**

***Keywords*—Video Streaming, Multicameras systems, Object Tracking.**

I. Introduction

In recent years, the issue of tracking the detected object in video captured by multiple is quite researched problem mainly for security systems sectors. Problems consisting of the detection of the object in diverse scenes of different colors from different angles have been the subject of several studies. However, the recent studies aren't focused to interconnection object detection and object tracking technologies with streaming technologies for the purpose of switching streaming source from several cameras in real time. Nowadays, there is a big demand on high quality video in surveillance system and also there is the demand for providing relevant and clear view on tracked object. New streaming technologies had to be designed to provide both quality and transfer time to be appropriate in surveillance systems. Tracking mechanisms must be effective and optimized for detecting object position in real time.

The Goal of this paper is to bring the approach, that interconnects streaming and object tracking technologies to solution, which would evaluate the position and size of tracked object for each field of view of each camera. Based on this evaluation, the solution would make dynamically switching of appropriate video source for resulting output stream accessible from web environment.

II. Streaming Technologies

Streaming technologies are designed for the transmission of multimedia content over a computer network in form of broadcast audio, video and additional content in the direction from the provider to the recipient. In this work this technology can be used for the streaming of the video content from multiple cameras in real time

A. Streaming platforms and codecs

For one of the most widely used platform for web-based streaming applications it could consider Flash and HTML5. Alternative platforms such as QuickTime and SilverLight technology have suffered the downward trend in the market in recent years.

Flash video can be encoded using codecs: Sorenson Spark (H.263 variant), On2 VP6 and H.264 (MPEG-4 Part 10) with encapsulation in file formats like: FLV, F4V, MP4, M4V, etc. [8].

HTML5 video elements provide simpler and more uniform integration of multimedia content in Web pages. It also provides independence from another software or third-party extensions. Currently it is certain inconsistencies in the use of codecs and HTML5 streaming formats in different browsers. H.264 encapsulated in MP4 container is supported by current versions of Internet Explorer, Safari, Google Chrome. Theora codec encapsulated in OGG container or VP8 codec encapsulated in WebM container is supported by current versions of Mozilla Firefox, Google Chrome and Opera. In terms of compatibility HTML5 technology can make a Flash fallback, when codec or format for HTML5 video is not supported by the browser. Flash fallback ability calls the external Flash Player and multimedia content will be accessible without any restrictions also for HTML5 unsupported browser. It is possible thanks to fact that H.264/AVC is usable codec for Flash and also for HTML5 platforms.

B. Streaming servers and protocols

Existing streaming protocols for transmitting multimedia contents vary in implementation details, which divide them into two main categories: the protocols with principle of

pushing stream (PUSH) and protocols with the principle of pulling stream (PULL). PUSH protocols establish a connection between the server and the client, which is still maintained and packets would send to the client until the connection is not interrupted by expiration of limit or disruption on the client site. Best known PUSH protocols are RTSP, RTMP, RTP, 3GPP. The PULL protocols make the client as active element that builds a connection in the form of requests for streaming media content from a server. The most famous use of the PULL protocol is the progressive downloading of multimedia content, where the client sends a HTTP request to the server and the content begins to download as soon as the server responds to the request [1].

III. Object Tracking in Video

Object detection in an image is used for identification and classification of these objects and to determine its characteristics such as size, rotation as well as position. To recognize an object in an image is necessary to divide the image into segments, reduce less important parts of the image, or to extract important parts of the image. Reduction or extraction of elements in the image can be realized by applications preprocessing filters on the image or by the description of the image based on its symptoms.

In terms of tracking the object in the video is important to realize the detection of the object on series of selected graduated frames of video. In each processed frame with successful detection of the object, the position of the object related to the shooting scene is calculated. The trajectory of the movement of object is defined by a series of information about object position in time, whether in the form of a sequence of points with coordinates or in form of the function.

Typical object tracking system consists of three components: a representation of the object, dynamic model and search engine. The object can be represented either holistic descriptors such as color histogram or the value of the brightness level of a pixel, or it may be represented by the local descriptors such as local histogram or color information. The dynamic model is typically used to simplify the computational complexity in tracking of the object. The search engine is used to optimize the object tracking and it can use deterministic and also stochastic methods. An important element of tracking methods is the motion model, which can represent for example the translational motion, the similarity transformation or the affine transformation [2].

A. *Real-time object tracking methods*

The current modern methods used in tracking objects in real time are: IVT [3], VRT [4], FRAG [5], boost [6], SEMI [7], BeSemiT [8], L1T [9], MIL [10], RTD [11] and the method of detection and tracking of learning ability (TLD) [12], which is shown as the most reliable and effective in the comparison of tracking methods implemented on various kind of videos with various kind of tracked object.

B. *TLD algorithm*

TLD (Tracking-Learning-Detection) algorithm is developed to detect and track objects in the video in real time. The object of interest is defined by the initialization of the square in the first frame, which serves as a template for detection. TLD simultaneously tracks the object, learns its properties in the following frames and detects and verifies the occurrence of the object in the image. The result is tracking of object in real time, where the accuracy and reliability of the tracking is improving related to time due to learning the new features of the observed object such as changing the position, size, rotation or brightness (Fig. 1).

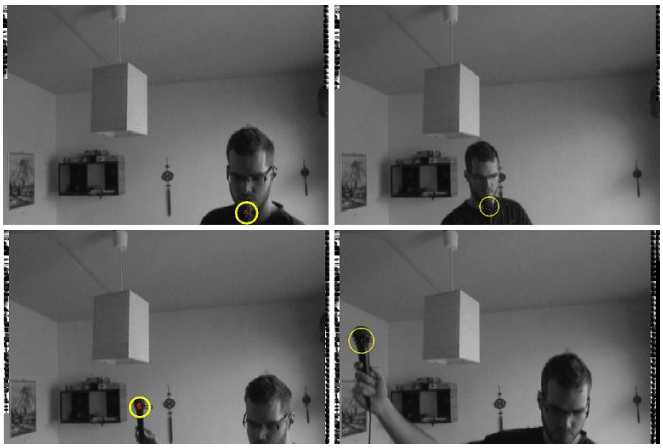


Fig. 1. Testing of tracking the microphone with the implementation of TLD algorithm in Matlab environment with using OpenCV library.

Tracking algorithm of TLD is based on the step of monitoring recursive tracking in forward and backward direction, which is performed by Lucas-Canade algorithm in both directions with calculating of median at the end. Detection of the TLD uses dispersion filter, file classifier and nearest neighbor classifiers. Learning in TLD is realized by PN learning, where the data can be classified with yielding the structure in the form of path of object and application of the positive (P) limitations and then the negative (N) limitation. Next stage is based on generating the new data and updating the classifier object. Segments of the image type P represent objects with a high probability of correlation with the template and the segments of image type N vice versa with low correlation [12,13].

IV. Design of Solution

Multi-cameras surveillance system should ideally cover a continuously monitored space and fields of view from nodes should overlap. When it comes to tracking objects that are viewed from different cameras with a different rotation, shape or texture, it is important the angle setting of each camera in Multi-cameras system. Ideally, it is expected that the proposed system contains N cameras, which are equally calibrated and spaced with spacing, in which the trajectory of a moving object was covered in all fields of view (FOV) of cameras that partially overlap.

For the purpose of processing information about the properties of the tracked object in the each camera frame at time t it is important to know its current position with related x and y coordinates and dimension of tracked object in the image with height h and width w and also the values reliability of detection d. The current state of the tracked object at time t for node n of Multi-camera system can be represented by vector (1):

$$O_n^t = [x, y, w, h, d] \quad (1)$$

The values contained in O_n^t will be used to calculate metrics (Fig. 2), which will be used for real time comparison between nodes in Multi-camera system.

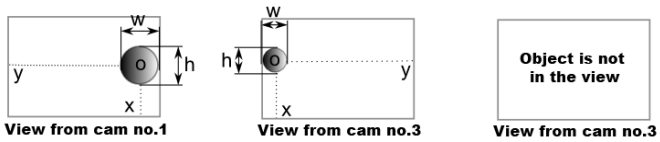


Fig. 2. Extracting parameters of state vector of tracked object from actual video frame from each node of multi-cameras system.

To optimize the performance of the tracking system it should be introduced frame detection filter that will collect the frame of video with step. This means that the detection mechanism treated with each i^{th} frame of a sequence. If consider the input video with frame rate f_s, the time between picking frame(2) for detection will be:

$$t_s = \frac{s}{f_s} \tag{2}$$

When considering the standard frame rate 25 frames per second with the selected step *s=5*, the image from frame is taken to object detection from a video every 200ms, among which it is necessary to process the image and detect the object to obtain information related to vector O_n^t. When it comes to capturing of fast moving objects in front of the camera with a higher frame rate, it should be used t_s values in tens of milliseconds, which can be depending on the resolution of an image a problem in terms of optimizing the performance of the detection or tracking mechanism.

For purposes of comparison between different nodes of the multi-cameras system and for evaluation of the video sources based on positional and dimensional characteristics of these objects, it is necessary to introduce an evaluation metric for evaluate each streaming source separately. The proposed metrics could consist of three components: a primary component related to position, the secondary component related to size and the tertiary component related to reliability of detection of the tracked object.

The primary positional component of the proposed metrics should be based on the current position of the anchor point of the tracked object in the field of view. The video frame needs to be divided into zones, which will represent a numerical rating. (Fig. 3)

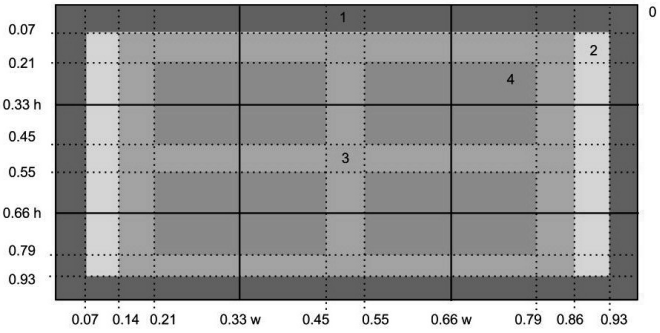


Fig. 3. Design of zones in video frame with values for evaluating the primary part of proposed metric related to position of tracked object in real time.

Final metric (3) of streaming source *n* at time *t* consisting of the components of M_p^t (positional component of metric), M_v^t (dimensional component), M_d^t (detection reliability component) could be expressed in the final form as:

$$M_n^t = (M_p^t + M_v^t) * M_d^t \tag{3}$$

Where M_p^t is an integer indicating the number of the zone of detection the object at time *t*, M_v^t component (4) is expressed in decimal form, where $M_v^t \leq 1$ is the aspect ratio of the observed object width at time *t* (w_o) and whole image width (w_z):

$$M_v^t = \frac{w_o}{w_z} \tag{4}$$

Component M_d^t indicates reliability of object detection in decimal form $M_d^t \leq 1$, which defines the percentage of similarity of tracked object with template defined to detect.

Based on the evaluation with proposed metric from *n* streaming sources obtained at time *t*, it could be defined the vector of evaluation metric (5) for all nodes in the system:

$$H_t = [M_1^t, M_2^t, \dots M_n^t] \tag{5}$$

After calculating the maximum $max_{0<k\leq n}\, M_k^t$ from *n* nodes at time *t*, it will be obtained the index of source with best evaluation metric in actual time, which determines which video source should be broadcasted in same time.

Streaming control mechanism for switching the streaming source based on actual highest evaluation metric could work directly with the mechanism for streaming playback on client side, or it could work via streaming server (Fig. 4).

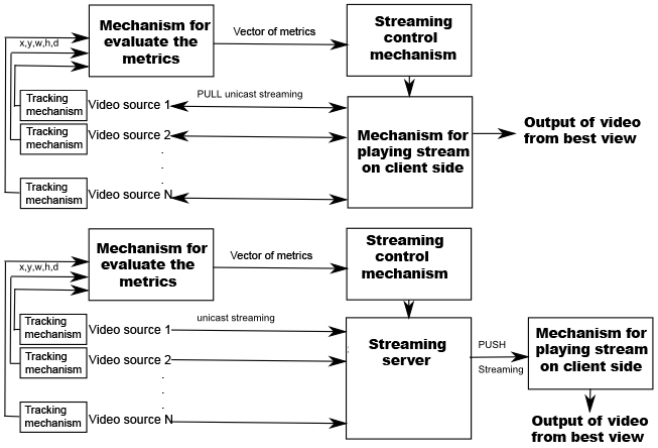


Fig. 4. Two concepts of solution for switching stream based on actual evaluation metric from N video sources.

In the first case, the player on client side connects with PULL streaming method to the address of the video source with the actual highest calculated metric and it starts to download the data stream from source. In the second case, the streaming server redistributes the resulting video source content with the highest evaluations to the playback mechanism on client side by PUSH method. Changing the source of output broadcast in both cases is related to changing values for defining the order of identification of the source.

V. Conclusion

Common multi-cameras tracking systems usually indicate the tracked object with highlighted graphics across all outputs

from cameras or they display position of the object in a global view on the map of the monitored area. Responsible person is therefore forced to watch the output from all cameras and he must visually search for the sign of tracked object on multiple windows with video stream or he is able to see only graphical representation of tracked object in global view on the area map, where the detail and characteristic of the object could be hidden from real view. This is the reason, why it is useful to design the solution for transferring the video from multiple cameras using streaming technology with displaying output just from one video camera, which detects the tracked object in a scene with the best positional results in actual time.

Using the proposed solution could be large scoped. It could be usable in surveillance systems, sports events, transport dispatching, educational conferences or monitoring of the production line and so on. The system could be applicable to a system of IP cameras in the network, to mobile devices or directly connected devices such as HD camcorders and video mixing panels. From an implementation point of view it could work for example with using technology OpenCV and OpenTLD for the detection and tracking mechanism, Wowza streaming server for streaming and SQL database technology for storing, processing and writing results of evaluation metrics of various video sources. Communicating between components in form of sending information about evaluation vectors of metrics could therefore work through the database, or it would be interesting to integrate this information directly in the header to used streaming transport protocol, what would bring better optimization and timing.

Acknowledgment

Research described in the paper was done by cooperation with Computer Networks Laboratory at DCI FEEI TU in Košice. This work is the result of the Project implementation: Competency Centre for Knowledge technologies applied in Innovation of Production Systems in Industry and Services, ITMS: 26220220155, supported by the Research & Development Operational Programme funded by the ERDF.

References

[1] BEGEN A. – AKGUL T. – BAUGHER M.: Watching Video over the Web, Part I, Streaming Protocols, Cisco, 2011

[2] WANG Q. a kol.: An Experimental Comparison of Online Object Tracking Algorithms, Proceedings of SPIE: Image and Signal Processing Track, 2011

[3] ROSS D. a kol.: Incremental learning for robust visual tracking, International Journal of Computer Vision 77(1-3), s. 125–141, 2008

[4] COLLINS R. – LIU T. – LEORDEANU M.: Online selection of discriminative tracking features, IEEE Transactions on Pattern Analysis and Machine Intelligence 27, s. 1631–1643, 2005.

[5] ADAM A. – RIVLIN E. – SHIMSHONI I.: Robust fragments-based tracking using the integral histogram, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, s. 798–805, 2006.

[6] GRABNER H. – BISCHOF H.: On-line boosting and vision, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, s. 260–267, 2006.

[7] GRABNER H. – LEISTNER C. – BISCHOF H.: Semi-supervised on-line boosting for robust tracking, Proceedings of European Conference on Computer Vision, s. 234–247, 2008.

[8] STALDER S. – GRABNER H. - VAN GOOL A.: Beyond semi-supervised tracking: Tracking should be as simple as detection, but not simpler than recognition, Proceedings of IEEE Workshop on Online Learning for Computer Vision,2009.

[9] MEI X. - LING H.: Robust visual tracking using l1 minimization, Proceedings of the IEEE International Conference on Computer Vision, s. 1436–1443, 2009.

[10] BABENKO B. – YANG M. – BELONGIE S.: Visual tracking with online multiple instance learning, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, s. 983–990, 2009

[11] KWON J. - LEE K.: Visual tracking decomposition, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 1269–1276, 2010.

[12] KALAL Z. – MATAS J. – Mikolajczyk K.: P-N learning: Bootstrapping binary classifiers by structural constraints, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 49–56, 2010.

[13] KALAL Z. - MIKOLAJCZYK K. - MATAS J.: Forward-Backward Error: Automatic Detection of Tracking Failures, International Conference on Pattern Recognition, Istanbul, Turkey, ICPR 2010

Teleoperation control of humanoid robot NAO with use of middleware system ROS

[1]*Marek Čopjak (1st year)*, [2]*Marek Čajkovský (2st year)*

Supervisor: [3]*Martin TOMAŠEK*

[1,2,3]Dept. Of Computers and Informatics, FEI TU of Košice, Slovak Republic
[1] marek.copjak@tuke.sk, [2]marek.cajkovsky@tuke.sk, [3]martin.tomasek@tuke.sk

***Abstract* — this paper deals with the field of teleoperation control of humanoid robot NAO, with use of middleware Robot Operating System (ROS). First it introduces the field of the humanoid robots and especially describes humanoid robot NAO. In the second part it introduces robotic middleware systems Robotics Technology Middleware (RTM) and special attention is paid to the Robot Operating System (ROS). Experiment with teleoperation control of NAO humanoid robot with use of ROS middleware is also described.**

Keywords: NAO, teleoperation, humanoid robot, ROS, robot operating system, RTM, Robotics Technology Middleware

I. INTRODUCTION

Traditional view of robots is a view of manipulators working in factories in environments that are tailored to the robots and are dehumanized. Robots are in this view separated from human everyday life. Nowadays, robotics is shifting from this point of view to the new generation of robots that are mobile, autonomous and are cooperating with humans in environments that are tailored to the human needs. To meet this demand, tens of humanoid robots have been developed in recent years. Humanoid robotics can be defined as the part of the robotics that focuses on construction of robots that emulates some subsets of the physical, cognitive and social dimensions of the humans.

Some projects are focused on cognitive and social aspects of humans, including projects of robots Cog and Kismet [1]. Other projects are focused on communication abilities of the human with projects including Memoni produced by Tomy and Tama, which is product of Matsushita.

Humanoid robots that are complex and are modeling aspects of human body can by divided according to their size into two groups. Humanoid robots that are taller than 120 cm are ARNE (Anthropomorphic Robot of the New Era). Male version is in white-blue design and is called ARNE, white-red is female version and is called ARNEA. GuRoo is the product of Mobile Robotics Laboratory at School of Information Technology and Electrical Engineering, University of Queensland. H7 is humanoid robot developed in Johou Systems Kougaku (JSK) Laboratory, Japan. Asimo (Advanced Step in Innovative Mobility) is the product of Honda and is known as the most developed humanoid robot (Fig. 1). Asimo is the offspring of the decades of development of the humanoid robots in Honda and its predecessors were called E0 to E6 and P1 to P3 [2]. HRP (Human Robotics Platform) is a series of humanoid robots developed by Kawada Industries [3]. HUBO (KHR-3) is the product of Korea Advanced Institute of Science and Technology (KAIST) and is offspring of the KHR series of humanoid robots.

Humanoid robots are emulating legs of humans, but there are projects that are using wheeled chassis instead, including MS DanceR (Mobile Smart Dance Robot) which is developed at the Tohoku University, Japan and Wakamaru which is the product of Mitsubishi Heavy Industries.

Figure 1. Asimo (Advanced Step in Innovative Mobility) humanoid robot developed by Honda.

Mid-sized and smaller robots, that are smaller than 75 cm are HOAP (Humanoid for Open Architecture Platform) as the series of humanoid robots developed by Fujitsu. HR6 is developed by Dr. Robot Corporation.

Small humanoid robot KHR is the product of Kondo Corporation and represents affordable humanoid robot that can be purchased on the mass market. Nuvo is the product of ZMP Corporation. PINO is the robot developed by Kitano Symbiotic. QRIO (Quest for cuRIOsity) is the humanoid robot developed by SONY

and is offspring of the SDR (Sony Dream Robot) series of humanoid robots. NAO is the product of the Aldebaran Robotics [4].

II. NAO HUMANOID ROBOT

In our experiments we are using product of France corporation Aldebaran robotics the humanoid robots NAO (Fig. 2). NAO is with 58 cm medium-sized humanoid robot, and its weight is 4.3 kg.

NAO has 2 speakers on the both sides of the head, and has the ability to synthesize voice and can speak in french and english language. It can reproduce also music in the wav and mp3 formats. Four microphones are implemented in the head of robot as the sound sensors for voice recognition and analysis of the system. Robot is able to recognize several voice commands and it can be trained for other voice commands recognition.

Figure 2. Aldebaran robotics humanoid robot NAO H25.

Two CMOS cameras with 640 x 480 pixels and with 30 frames per second are implemented in the head. One of them is in the forehead of the robot and this camera is used for acquisition of visual information from the broader environment of the robot. Another camera is situated in the mouth of the robot and is used for acquisition of visual information from the near environment of the robot.

NAO is able to recognize several faces of people and can recognize simple and complex objects, for example ball. Special modules for computer vision can be designed with use of OpenCV (Open Computer Vision) interface.

NAO is able to electronically communicate with use of several interfaces. IRDA transmitter/receiver placed in his eyes serves for control of remote controlled devices in the environment. NAO can connect to the local area network via Ethernet or Wi-fi (IEEE 802.11g). It has also USB connectivity.

Computing capacity of the robot is powered by X86 AMD GEODE 500 MHz CPU, 256 MB SDRAM operating memory and 2 GB of flash memory. ARM7 microcontroller with 60 MHz frequency is located in the torso of the robot and is managing all microcontrollers of actuators Microchip 16 bit dsPICS. Two RS485 busses are connecting the ARM7 microcontroller with dsPICS modules, one bus in the upper part of the robot's body and one bus in the lower part of the robot's body.

NAO has AC 90-230 V / DC 24 V power supply and robot is able to work for 90 minutes on the single charge of its batteries.

NAO has 25 degrees of freedom (DOF), 2 DOF in the head, 2x5 DOF in the arms, 2x1 DOF in palms, 1 DOF in the pelvis and 2x5 DOF in the legs. Detailed information is shown in the table 1.

Operating system of the robot is Embedded Linux which is the own distribution of Linux operating system and can be programmed in several computer languages, including C, C++, URBI and Python. Aldebaran Robotics also supplies graphical programming software Choreographe. With this software NAO can be visually programmed by interconnection of preprogrammed blocks of activities (Tab. 1).

TABLE I.
SERVOMOTORS USED IN JOINTS OF THE ROBOT

Part	Identificator	Movement	Angle range
Head	HeadYaw	Head joint twist (Z)	-120/120
	HeadPitch	Head joint front back (Y)	-39/30
Left arm	LShoulderPitch	Left shoulder joint front & back (Y)	-120/120
	LshoulderRoll	Left shoulder joint right & left (Z)	0/95
	LElbowRoll	Left shoulder joint twist (X)	-90/0
	LElbowYaw	Left elbow joint (Z)	-120/120
	LWristYaw	Left wrist joint twist (X)	-105/105
	LHand	Left hand	open/close
Right arm	RShoulderPitch	Right shoulder joint front&back (Y)	-120/120
	RShoulderRoll	Right shoulder joint right & left (Z)	-95/0
	RElbowRoll	Right shoulder joint twist (X)	0/90
	RElbowYaw	Right elbow joint (Z)	-120/120
	RWristYaw	Right wrist joint twist (X)	-105/105
	RHand	Right hand	open/close
Left leg	LHipYawPitch	Left hip joint twist (Z45°)	-44/68
	LHipPitch	Left hip joint front & back (Y)	-104.5/28.5
	LHipRoll	Left hip joint right & left (X)	-25/45
	LKneePitch	Left knee joint (Y)	-5/125
	LAnklePitch	Left ankle joint front & back (Y)	-70.5/54
	LAnkleRoll	Left ankle joint right & left (X)	-45/25
Right leg	RHipYawPitch	Right hip joint twist (Z45°)	-68/44
	RHipPitch	Right hip joint front & back (Y)	-104.5/28.5
	RHipRoll	Right hip joint right & left (X)	-25/45
	RKneePitch	Right knee joint (Y)	-5/125
	RAnklePitch	Right ankle joint front & back (Y)	-70.5/54
	RAnkleRoll	Right ankle joint right & left (X)	-45/25

III. RTM AND ROS MIDDLEWARE

Robotic middleware is the software, which is creating interconnection between different hardware and operating system platforms and applications that are running on those platforms on the one side and embedded system [5][6] of the robot on the other side. Architecture of robotic middleware (Fig. 3) was developed with the aim to ensure interoperability in distributed architectures [7][8].

Robotic middleware includes Microsoft Robotics Development Studio (MRDS) [9], OLE for Process Control (OPC), Open Platform for Robotic Services (OPRoS) [10][11], Open Robot/Resource Interface Network (ORiN) and Robot Software Communications Architecture (RSCA).

Robotics Technology Middleware (RTM)

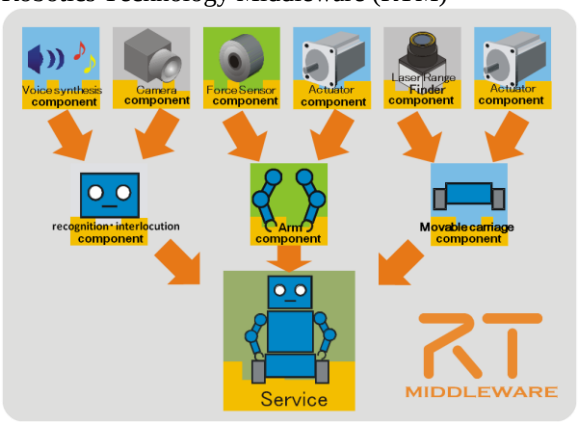


Figure 3. Robotic system created by RT-Middleware and RT-Component [15]

RT-Middleware is a project that includes a software framework and also platform, which allows to build robotic systems by combining smaller functional robotic parts, which are called components. Component can provide comprehensive functionality, which can be a device as a servomotor, sensor and camera, or a combination of devices such as a robotic arm or mobile robotic platform. The component may be more tied to specific hardware, but can also be simply software, such as image processing components, control components (comprising management and control algorithms). In the RT-middleware such component is called RT-Component, abbreviated RTC. Each component has its own interface, called port for data exchange and for communication with other components. The robot is built from the linked ports of components. [12][13][14] [15].

ROS is a framework used for creation of robotic applications, which functions as an operating system. It is designed primarily for UNIX based operating systems, but some of its components are capable of running under different types of operating systems including Windows. [16][17][18][19].

ROS was originally designed for solving specific tasks which can be found in the process of creation of complex service robots as the part of STAIR (STanford AI Robot) project at Stanford University and Personal Robots Program which is developed by Willow Garage, but the resulting architecture is much more generic as the mentioned domains.

Systems built on a ROS basis consist of a number of processes or different kinds of computers connected to each other via peer-to-peer topology. Nowadays ROS supports four very different programming languages: C++, Python, Octave and LISP. Support for other languages is in different state of finalization for each language. ROS uses simple, language-neutral interface definitions of the IDL language, which describes the messages sent between the modules.

The source code generator for each individual and supported programming language generates native implementations, which behave as native objects and are automatically serialized and reconstructed by ROS based on how the messages are sent and received. The final result is a language neutral schema for message handling, which can contain many languages mixed and enclosed as needed.

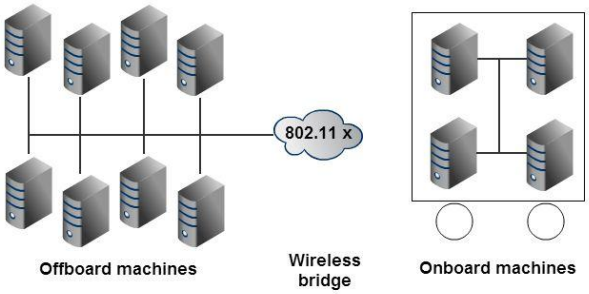


Figure 4. Typical network configuration of ROS middleware [16]

To handle the system complexity of ROS as well as possible, it was decided to use a micro-kernel design, where a lot of small tools are used to create and run different ROS components. These tools fulfill many tasks, such as navigation in the source code tree, getting and setting configuration parameters, peer-to-peer connection visualization in the topology and so on. The whole ROS source code is available to the public. It is distributed under the agreements of the BSD license, which allows the development of commercial and non-commercial projects.

IV. EXPERIMENTS

In our experiments aimed on teleoperation control (Fig. 4) of humanoid robot we have been using a humanoid robot NAO. As the first step we have been evaluating different robotic middleware and we chose Robot Operating System (ROS), because of the large number of components oriented on NAO robot with reach documentation, that are available also with source code under the BSD license. In the experiment we modified one of ROS components nao_teleop. Modification of ROS component consisted of the implementation of upper limbs control. On modified component we connected through the host computer gamepad PS3 for control of the robot moves and component is connected to the roscore, which is in the center of ROS middleware and connects all components. In experiment another two standard ROS components nao_driver and nao_vision without any modification were used. First component nao_driver was used for direct robot motion control; second component nao_vision (Fig. 5) was used for processing of video stream acquired from NAO visual subsystem and for transmission of this video stream to the host computer, where NAO was teleoperated by the human.

In our teleoperation experiments robot NAO was wirelessly connected through the Wi-Fi connection to the computer network together with server computer and host computer and different configurations of software components distribution were evaluated, with

modifications, when nao_driver was running on the NAO hardware or along with other ROS components and roscore were running alternatively on the server or on the host computer.

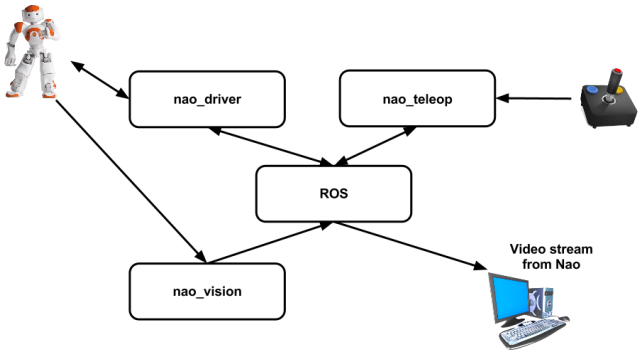


Figure 5. Schematic representation of teleoperation control of NAO via ROS middleware

V. CONCLUSION

This paper deals with the field of teleoperation control of humanoid robot NAO, with use of middleware Robot Operating System (ROS).

First it introduces the field of the humanoid robots and especially describes humanoid robot NAO. In the second part it introduces robotic middleware systems Microsoft Robotics Development Studio (MRDS), OLE for Process Control (OPC), Open Platform for Robotic Services (OPRoS), Open Robot/Resource Interface Network (ORiN), Robot Software Communications Architecture (RSCA) and Robotics Technology Middleware (RTM) and special attention is paid to the Robot Operating System (ROS). Experiment with teleoperation control of NAO humanoid robot with use of ROS middleware is also described.

In the future research attention will be paid to the refinement of the reliability of communication channel between human operator and robot, to ensure continuity of robot control and mechanisms will be proposed to deal with the situations when communication channels will be interrupted.

ACKNOWLEDGMENT

This work was supported by the Slovak Research and Development Agency under the contract No. APVV-0073-07 and No. APVV-0008-10. Projects are being solved at the Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics, Technical University of Košice.

REFERENCES

[1] C. Breazeal, "Designing Sociable Robots" A Bradford Book, August 20, 2004, ISBN-13: 978-0262524315, pp. 281.

[2] Y. Sakagami, R. Watanabe, C. Aoyama, S. Matsunaga, N. Higaki, and K. Fujimura, "The intelligent ASIMO: system overview and integration," IEEE/RSJ Int. Conference on Intelligent Robot and System, vol. 3, pp. 2478–2483, 2002.

[3] K. Akachi, K. Kaneko, N. Kanehira, S. Ota, G. Miyamori, M. Hirata, S. Kajita, and F. Kanehiro, "Development of humanoid robot HRP- 3P," 5th IEEE-RAS International Conference on Humanoid Robots, pp. 50–55, 2005.

[4] D. Gouaillier, V. Hugel, P. Blazevic, Ch. Kilner, J. Monceaux, p. Lafourcade, B. Marnier, J. Serre, and B. Maisonnier, "Mechatronic design of NAO humanoid". *In Proceedings of the Open-Source Software Workshop of the International Conference on Robotics and Automation.*, Kobe, May 17th, 2009.

[5] J. Kopják, and J. Kovács, "Timed cooperative multitask for tiny real-time embedded systems". In: 10th Jubilee International Symposium on Applied Machine Intelligence and Informatics (SAMI 2012), Herl'any, Slovakia, pp. 377-382, ISBN 978-1-4577-0197-9, 2012.

[6] J. Kopják, and J. Kovács, "Event-driven control program models running on embedded systems". In: 6th IEEE International Symposium on Applied Computational Intelligence and Informatics, Timişoara, Romania, pp. 323-326, ISBN 978-1-4244-9107-0, 2011.

[7] L. Vokorokos, A. Baláž, N. Ádám, S. Petrík, "Dataflow distributed database systems", In: DAAAM International, Wien, pp.373-374, ISBN 3901509461, 2005.

[8] N. Ádám, L. Vokorokos, B. Madoš, "Modeling of intelligent calculations in cubic network of workstations", IN: SAMI 2009. - s.l. : IEEE, 2009 S. 147-151. - ISBN 9781424438020

[9] D. S. Michal, and L. Etzkorn, "A Comparison of Player / Stage / Gazebo and Microsoft Robotics Developer Studio,", 49th Association for computing machinery, ACM Southeast Conference, March 24-26, 2011, Kennesaw, GA, USA. Copyright 2011 ACM 978-1-4503- 0686-7/11/03.

[10] B. Song, S. Jung, C. Jang, and S.Kim. "An Introduction to Robot Component Model for OPROS", In Proceedings of Simulation, modeling and programming for autonomous robots SIMPAR 2008, Venice, Italy, Nov. 3.-7. 2008, pp. 592-603, ISBN 978-3-540-89075-1.

[11] The Open Platform for Robotic Services (OPRoS), official homepage of the OPRoS project, http://www.opros.or.kr/.

[12] N. Ando, T. Suehiro, K. Kitagaki, T. Kotoku, and W. Yoon. RTComponent Object Model in RTMiddleware - Distributed Component Middleware for RT (Robot Technology). In IEEE Int. Symposium on Computational Intelligence in Robotics and Automation (CIRA), 2005.

[13] Y. Hada, S. Jia, K. Takase, H. Gakuhari, and T. Ohnishi, "Development of Home Robot Integration System Based on Robot Technology Middleware,", The 36th International Symposium on Robotics (ISR 2005), Tokyo, Japan, Nov. 29 – Dec. 1., 2005.

[14] G. Hwang, N. Ando, H. Hashimoto, " RT-Middleware Network Framework for Single-Master Multi-Slave Tele-micromanipulation System", Proc. Of 36th International Symposium on Robotics (ISR 2005), Tokyo, Japan, Nov. 29 – Dec. 1., 2005.

[15] S. Cocuľa, "RT-Middleware pre robota NAO", Diplomová práca, Košice: Technical University of Košice, Department of computers and informatics 2012, 87 s.

[16] M. Quigley, B. Gerkey, K. Conley, J. Faust, T. Foote, J. Leibs, E. Berger, R. Wheeler, and A. Ng, "ROS: an open-source Robot Operating System," *In Proceedings of the Open-Source Software Workshop of the International Conference on Robotics and Automation.*, Kobe, May 17th, 2009.

[17] C. Crick, G. Jay, S. Osentoski, B. Pitzer, and O. C. Jenkins. Rosbridge: "Ros for non-ros users," *In Proceedings of the 15th International Symposium on Robotics Research*, 2011.

[18] S. Osentoski, G. Jay, C. Crick, B. Pitzer, C. DuHadaway, and O. C. Jenkins. "Robots as web services: reproducible experimentation and application development using rosjs," *In Proceedings of the 2011 IEEE International Conference on Robotics and Automation*, 2011.

[19] J. García-Nieto, E. Alba, and F. Chicano, "Using Metaheuristic Algorithms Remotely via ROS," Genetic and Evolutionary Computation Conference *GECCO'07*, July 7–11, 2007, University college London, London, England, United Kingdom. ACM 978-1-59593-697-4/07/0007.L. Vokorokos, A. Baláž, N. Ádám, S. Petrík, "Dataflow distributed database systems", In: DAAAM International, Wien, pp.373-374, ISBN 3901509461, 2005.

The Importance of Embodiment for Intuitive Human-Computer Interaction

[1]*Peter TAKÁČ,* [2]*Mária VIRČÍKOVÁ (3rd)*
Supervisor:[3]*Peter SINČÁK*

[1,2,3] Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

[1]takac.p.x@gmail.sk, [2]maria.vircikova@tuke.sk, [3]peter.sincak@tuke.sk

***Abstract*— Human—Computer Interaction(HCI) is one of the most spelled words, when it comes to new product development, which will be used by public. However, not everybody is aware of what does it take to develop a successful and widely used product and what are the main goals they should take in consideration. The paper focuses on the contribution of the embodiment—a humanoid robot—for intuitive interfaces in an application of a conversational agent. We also discuss the newest problems and fields of study of HCI, what does it do to accomplish their goals and what is the course of their work.**

***Keywords*—embodiment, humanoid robots, human-computer interaction, intelligent interfaces.**

I. INTRODUCTION

"We need not take human-human conversation
as the gold standard for conversational exchanges.
If one had a perfect simulation of a human conversant,
then it would be human-human conversation
and not human computer conversation
with its sometimes odd but pertinent properties."
Colby in [17]

Our physical bodies play a central role in shaping human experience in the world, understanding of the world, and interactions in the world. For example, there is a qualitative difference between face-to-face conversation and other forms of human-human communication [1]. When it comes to human-computer interfaces, we argue that embodied interface agents can provide a qualitative advantage over non-embodied interfaces, if the bodies are used in ways that leverage knowledge of human communicative behavior.

A study by [3] proves that the qualitative difference in the interaction such as a conversation, is not just that we enjoy looking at humans more than at computer screens but also that the human body enables the use of certain communication protocols in face-to-face conversation which provide for a more rich and robust channel of communication. The use of gaze, gesture, intonation, and body posture play an essential role in the proper execution of many conversational functions such as conversation initiation and termination, turn—taking, interruption handling, feedback and error correction and these kinds of behaviors enable the exchange of multiple levels of information in real time.

We adopt these functions from human—human communication and implement into the embodied conversational character. Usually, this term is used by researchers who focus on HCI with non-verbal behavior, mainly in the form of 3D animations and verbal behavior (speech synthesis and speech recognition systems). In this paper we refer to the embodiment in the form of a humanoid robot Nao—a programmable platform with 25 degrees of freedom. It features four microphones (for voice recognition and sound localization), two speakers (for text-to-speech synthesis) and two HD cameras (for computer vision, including facial and shape recognition).

Our outgoing research studies the contribution of this type of embodiment which brings new possibilities to the construction of a conversational agent—implementation of chatbot, like famous Eliza (implemented by Joseph Weizenbaum in 1966) or ALICE (Artificial Linguistic Internet Computer Entity implemented by Dr. Richard S. Wallace in 1995). We agree with Wilks [16] that the need of conversational agents has become acute with the widespread use of personal machines with the wish to communicate and the desire of their makers to provide natural language interfaces.

The paper is organized as follows: we first review field of human—computer interaction in general. We then discuss the embodiment using humanoid platform for the conversational human—computer interface. Finally, we describe our ongoing research program to develop embodied interface agents.

II. ROOTS AND FUTURE OF HUMAN-COMPUTER INTERACTION

Human-computer interaction (HCI) is an area of research and practice that emerged in the early 1980s, initially as a specialty area in computer science embracing cognitive science and human factors engineering [4]. HCI has expanded rapidly and steadily for three decades, attracting professionals from many other disciplines and incorporating diverse concepts and approaches. To a considerable extent, HCI now aggregates a collection of semi-autonomous fields of research and practice in human-centered informatics.

The main goal of HCI is usability and simplification of interaction between human and computer. According to [12], products of HCI regarding to usability should satisfy the three

U rule, which is Useful, Usable and Used. An interactive system is useful, when it focuses on the main problem and is not distracted by side tasks. Usable and used are two rules important for the end user of the system. Usability is important, if we want the user to learn to work in the new environment in the shortest time, so it should be easy to use, work in way natural to human, it also should be aware of errors and fix these errors. But previous two rules are for nothing, if the system will not be used. Usage of systems is another point of interest of HCI. This rule contains attractiveness, so the final system should be attractive for users, thus they will be motivated to use it. In modern day, usability often subsumes qualities like fun, well-being, collective efficacy, aesthetic tension, enhanced creativity, flow, support for human development, and others.

Simplification in HCI is another very important area of approach, especially if we want our products to be used by large variety of users regardless of skill level of users, that will use our products. The main task of simplification is to make products, that will be easy to learn and easy to use, i.e. we are trying to optimize for as many users as we can, while we observe every given restriction.

However, research in the field of usability and simplification does not stop and pushes the development towards new solutions of interactions.

HCI involves the study, planning, design and implementation of interactive systems for ordinary users. It is often regarded as the intersection of computer technologies and science, behavioral science, design and other fields of study such as psychology or sociology. In fact, HCI interferes almost every part of computer science, because there is always some sort of human connection and some sort of interaction between human and computer must be established. These are the main reasons why is the area of the research of HCI so wide and complex and makes no sense to regard HCI as a specialty of computer science.

There are two main fields of study in HCI, while it is an interaction between two actors human and computer. HCI benefits from supportive knowledge on both sides machine and human.

HCI draws from knowledge about humans like linguistics, psychology, biology, sociology etc. or tries to understand how we communicate with one another, human decision making, prediction and intuition, experience gaining and much more. HCI also tries to find corresponding algorithms or mechanisms, which will match the natural mechanisms of human side of interaction. This way we can make an imaginary cycle during development of new methods of interaction, technology, tools or environments, where on one side is the task, which is given be human needs and on the other side is the artifact, that not only responds, but also gives feedback about the possibilities of the interaction. That's why it is important to understand of the human processing.

On the machine side, techniques in computer graphics, operating systems, programming languages and development environments are relevant to HCI. The key reason for this is our requirement for computer to understand the task given and to respond according to the given information and to satisfy the user solving the task.

Current research in HCI is moving towards new possibilities of interaction. Tries to find new ways how can humans communicate with computers. Another research approach is to offer the user maximum freedom in customization of their products. This way we can determine the needs of the end-user community and we can collect knowledge about their domains, what will be useful source of information in development of new applications. To recover information from end-users we need an application or device, which will collect these data. That is why is HCI trying to understand embedded systems. Embedded systems allow computations automatize processes on almost every device, which gives us an edge for collection of data, because there are much more sources, than just an ordinary desktop. These devices are mainly mobile phones, tablets, but also washing machines or cars.

More futuristic approach is using augmented reality to communicate with users. This should allow user to communicate and collect information real-time. Resources of augmented reality are mostly mobile phones, tablets, GPS, but also Laptops. i.e. everything, that can process audio or video, which are the two most important elements of augmented reality. Augmented reality differs from virtual reality. As argued by [9][9], virtual reality replaces the real world with a simulated one, while augmentation is usually in real-time and in semantic context with environmental elements.

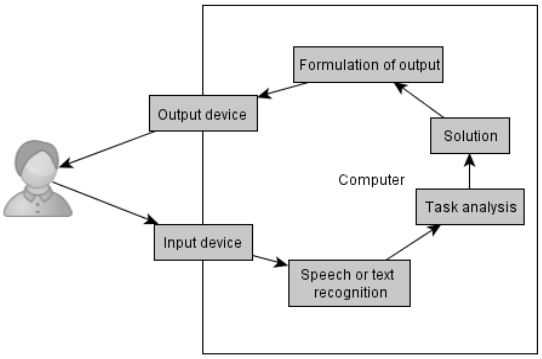


Fig. 1. Simplified human-computer interaction diagram, inspired by [12]. It is a computer side oriented interaction, where the user gives an input task using the computers input device (can be any other device or machine, which interferes with human), this could be speech, text or any other form of input, which is familiar to our system. Then the computer analyses the given problem using predefined algorithms. Output of these algorithms or mechanisms is usually the desired output, but it is in form of computer language so it is formulated into desired output (video, text, audio). And then it is processed to the user.

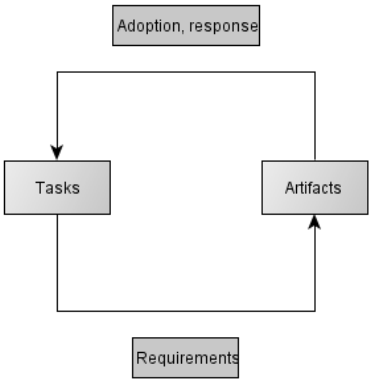


Fig. 2. Task-artifact cycle by [12]. Tasks match human processing and information given by the user. Requirements match the needed resources of the communication between human and computer. Artifacts are the computer processing algorithms or mechanisms. Adoption and response are the output of the computer by which we can rate the level of communication.

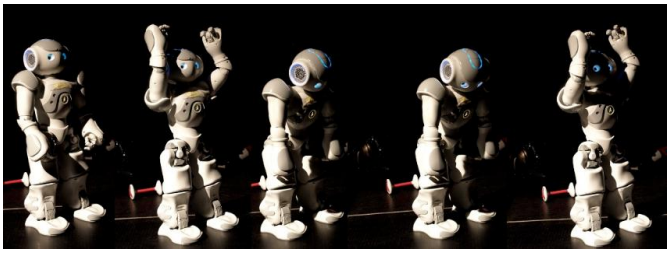


Fig. 3. Humanoid robot NAO. The embodiment can bring new possibilities for the implementation of chatbot, for example, include the non-verbal part of communication among others.

III. Embodiment For Conversational Agent

We use the humanoid robot NAO as an embodied chatbot. In our application, the platform represents the output mechanism. Using the built in text to speech function, we can easily convert the text based output of the chatbot to speech, which will be more acceptable output for the user. The input text can be also easily realized by sending a message to NAO via chatbox. Another approach is using a speech recognition mechanism which will convert normal human speech to text so NAO can react to it. The core of the research consists in comparing the human-computer interaction during the conversation using a conventional computer and using the humanoid robot. We believe that humanoid robots have capabilities to make the interaction more natural for the human user and thus, more effective.

A chatbot is usually an artificial person, which communicates with humans. The first was the ELIZA program by Weizenbaum in 1966 which emulated a psychotherapist. Conversation can be a text based (typed) conversation, a spoken conversation or even a non-verbal conversation. Chatbots can run on local computers and phones, but most of the time they are accessed through the internet. Chatbots are typically perceived as engaging software entity, which humans can talk to. Some chatbots use sophisticated natural language processing systems, but many simply scan for keywords within the input and pull a reply with the most matching keywords or the most similar wording pattern from a textual database.

Chatbots using a natural language processing system are also called artificial intelligence chatbots. Natural language processing is one of the most critical part of creating an artificial intelligence chatbot, while response of the chatbot depends on understanding the input, i.e. natural language understanding, and correct processing of natural language processing.

The major tasks of natural language processing are either helping the accomplishment of bigger and more complex tasks or research and find new real-world applications. Some of the tasks, that a natural language processing should be able to fulfill are recognition of speech or text input, where the most important part is machine translation, analysis of the input language, which includes linguistic analysis, if it is allowed then also determining emotional influences, information retrieval and extraction, last but not least question answering and formulation of output.

There is much work on embodied conversational agents like virtual humans capable of carrying on conversations with humans by both understanding and producing speech, hand gesture and facial expressions in games or online assistants. In humanoid robotics there is a need to give the ability of maintaining a conversation with humans especially in the area of social robotics. From the perspective of artificial intelligence, Brooks in [5]and [6] referred to embodiment as physical grounding, arguing that "to build a system that is intelligent it is necessary to have its representations grounded in the physical world ... [i.e.] it is necessary to connect it to the world via a set of sensors and actuators."

Barsalou [7] have addressed the notion of social embodiment by which they mean that states of the body, such as postures, arm movements, and facial expressions, arise during social interaction and play central roles in social information processing this aspect is addressed indirectly in their discussion of four types of "embodiment effects": First of all, perceived social stimuli don't just produce cognitive states, they produce bodily states as well. Secondly they argue that perceiving bodily states in others produces bodily mimicry in the self. Then, bodily states in the self produce affective states. The compatibility of bodily states and cognitive states modulates performance effectiveness.

Embodiment allows intuitive interactions. Giving the robot a similar body shape also makes interaction with humans more intuitive. Users can easily interpret the robots physical movements as the elements of non-verbal communication.

Communicative behaviors such as salutations and farewells, conversational turn-taking with interruptions, and describing objects using hand gestures are examples that all native speakers of a language already know how to perform and can thus be leveraged in an intelligent interface.

IV. Discussion

In the future we can expect much more complex interfaces, but on the other side more simple for end-user. Interfaces will be fully automatized and there will be no need of any other peripheral to control the computer or any other device – hands free control. Only peripherals for control will be our body. Algorithms will improve even further and we get to a whole new level of control over interactive computer systems. Also new ways of communication will appear like facial expressions and gestures, not only speech.

Some devices will work independently a will not require human control, but they will automatically adjust to human needs. We will take for granted systems, which interfaces will be fully and easily adjustable for every single and unique user, so it fits their needs perfectly.

Progress in recognizing emotions will move forward and we will maybe see systems able to recognize and react to human emotions correctly, what is a very difficult task if we take into count, that even humans communicating with each other have sometimes problems recognizing emotions of their communication partner.

But interaction between computer and human does not mean only understanding the human side of communication. If we want to close our communications circle we need to determine also computer side communication. In future there will be more powerful computers so computations will be extremely fast, not to mention the always ongoing miniaturization and cross-platform programming, which helps spreading technologies on new platforms like cars, dishwashers or even toilets. In theory we can implement HCI systems to almost every device, system or machine, which get in contact with humans.

We will continue to explore the nature of embodiment as it relates to the area of Human-Robot Interaction. As part of an

ongoing research program into socially assistive robotics, we intend to compare a physical robot to a non-embodied agent with a great application potential. We need to test the system in various environments with the highest priority in care of elder and disabled people-to entertain them and to assist them in daily routines.

REFERENCES

[1] S. R. Klemmer and B. Hartmann, "How Bodies Matter: Five Themes for Interaction Design," in *Carnegie Mellon University report*, 2006, pp. 140-149.

[2] E. Boyle, A. Anderson, A. Newlands, "The effects of visibility in a cooperative problem solving task", in *Language and Speech*, 37/1, 1994.

[3] J. Cassell, T. Bickmore, L. Campbell and H. Vilhjaâlmsson and H. Yan, "More than just a pretty face: conversational protocols and the affordances of embodiment," in *Knowledge-Based Systems*, 2001, pp. 55-64.

[4] J. M. Carroll, John M, "Human Computer Interaction (HCI)", in: Soegaard, Mads and Dam, Rikke Friis (eds.), *The Encyclopedia of Human-Computer Interaction*, 2nd Ed.. Aarhus, Denmark: The Interaction Design Foundation. Available online at http://www.interaction-design.org/encyclopedia/human_computer_interaction_hci.html, 2013.

[5] R. Brooks, "Elephants don't play chess", in *Robotics and Autonomous Systems*, 6(1-2), pp. 1-16, 1990.

[6] R. Brooks, "Intelligence Without Reason", in *Proc. of the Twelfth Intl. Joint Conf. on Artificial Intelligence*. San Mateo, CA: Academic Press, 1991.

[7] L. Barsalou, P. Niedenthal, A. Barbey and J. Ruppert, J, "Social embodiment", in B. Ross (Ed.), *The Psychology of Learning and Motivation*, Vol. 43. San Diego, CA: Academic Press, 2007

[8] K. Dautenhahn, B. Ogden, and T. Quick, "From embodied to socially embedded agents – implications for interaction-aware robots", in *Cognitive Systems Research*, 3(3):397–428, 2002.

[9] V. Kaptelinin, "Activity Theory", in Soegaard, Mads and Dam, Rikke Friis (eds.), *Encyclopedia of Human-Computer Interaction*, 2013.

[10] B. X. Chen, "If You're Not Seeing Data, You're Not Seeing", in *Wired*, published 25 August 2009. Available online at http://www.wired.com/gadgetlab/2009/08/augmented-reality/

[11] K. Maxwell, "Augmented Reality", in *Macmillan Dictionary Buzzword*. Available online at http://www.macmillandictionary.com/buzzword/entries/augmented-reality.html

[12] J. Lowgren, "Interaction Design", in Soegaard, Mads and Dam, Rikke Friis (eds.). *The Encyclopedia of Human-Computer Interaction*, 2nd Ed.". Aarhus, Denmark, 2013.

[13] T. Hewett, R. Baecker, S. Card, T. Carey,J. Gasen, M. Mantei, G. Perlman, G. Strong and W. Verplank, "ACM SIGCHI Curricula for Human-Computer Interaction", in *Report of the ACM Special Interest Group on Computer-Human Interaction (SIGCHI) Curriculum Development Group*. Available online at http://old.sigchi.org/cdg/

[14] A. Kerly, P.Hall, S.Bull, „Bringing chatbots into education: Towards natural language negotiation of open learner models", in *Knowledge-Based Systems*, Volume 20, Issue 2, March 2007, pp. 177-185, ISSN 0950-7051.

[15] W. Zadrozny, M. Budzikowska, J. Chai, and N. Kambhatla, "Natural language dialogue for personalized interaction", in *Communications of the ACM*, 43(8): pp. 116–120, 2000.

[16] Wilks, Y, "Preface", in Wilks, Y., editor, *Machine Conversations*, pp. 8-10. Kluwer, Boston/-Dordrecht/London, 1999.

[17] B. A. Shawar, and E. S. Atwell, "Chatbots: Are they really useful?", in *LDV-Forum*, 22, pp. 31–50, 2007.

Turn-based Games and Minmax Family of Algorithms

[1]*Peter SZABÓ (2nd year)*
Supervisor: [2]Kristína MACHOVÁ

[1,2]Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

[1]peter.szabo.2@tuke.sk, [2]kristina.machova@tuke.sk

***Abstract*—This paper is focused on the area of turn-based games and makes a review of minmax algorithms. It divide games into two categories, describes the basic terms and the use of a game tree. The minmax algorithm is used to describe basic tree search and it is used as a reference to all algorithms introduced in this article. Finally, the importance of the algorithm towards board games is concluded and the future research direction is outlined.**

***Keywords*—Game tree, Minmax, Turn-based games, State space**

I. Introduction

The games are part of our lives since childhood. We spend time with playing and most of us are still playing games today, although they are not the same as they was. In computer industry there is a rich history behind the games and even today we are writing a new lines. Since the growth of interactive web, social network and mobile platforms we are literally surrounded with infinite possibilities to play the game.

With increasing computational power humans attempted to let computers play games. Chess was one of the first games that received attention from science [1]. The first persons to describe a possible chess program were Shannon and Turing [2].

We can divide games into the two categories:

1. Real-time games
2. Turn-based games

In **real-time games**, game time progresses continuously according to the game clock. Players perform actions simultaneously as opposed to in sequential units or turns. Players must perform actions with the consideration that their opponents are actively working against them in real time, and may act at any moment. This introduces time management considerations and additional challenges (such as physical coordination in the case of video games).

In **turn-based games**, game flow is partitioned into well-defined and visible parts, called turns. A player of a turn-based game is allowed a period of analysis before committing to a game action, ensuring a separation between the game flow and the thinking process, which in turn presumably leads to better choices. Once every player has taken his or her turn, that round of play is over, and any special shared processing is done. This is followed by the next round of play. In games where the game flow unit is time, turns may represent such things as years, months, weeks or days.

Although the area of real-time games is very interesting [1] [3] and challenging, this publication and further work are focused on the area of turn-based games.

II. Turn-Based Games

The domain of the turn-based games is wide and belongs here areas such as reinforcement learning, general game play or game tree search. In this paper we have chosen the game tree search as a main area of the study.

The best known and most advanced algorithms for turn-based games are designed to work with two-player, zero-sum, perfect information games [1].

The following subtopics describes this basic terms in the area of turn-based games.

A. Number of Players

The board games that inspired turn-based AI algorithms almost all have two players. Most of the popular algorithms are therefore limited to two players in their most basic form. They can be adapted for use with larger numbers, but it is rare to find descriptions of the algorithms for more than two players.

In addition, most of the optimizations for these algorithms assume that there are only two players. While the basic algorithms are adaptable, most of the optimizations cannot be used as easily.

B. Zero-Sum Games and the Goal of the Game

In most board games the aim of the game is to win. As a player you can win if all your opponents lose. This is known as a **zero-sum game**: your win is your opponent's loss. If you scored 1 point for winning, then it would be equivalent to scoring −1 for losing. In a zero-sum game it does not matter if you try to win or if you try to make your opponent lose.

C. Perfect / Imperfect Information

In games like Chess and Reversi, both players know everything there is to know about the state of the game. They know the results of every move and what is the options for the next move. They know all this from the start of the game. This kind of game is called **perfect information**. Although you do not know which move your opponent will choose to make, you have complete knowledge of every move your opponent could possibly make and the effects it would have.

In a game such as Backgammon, there is a random element. You do not know in advance of your dice roll what moves you will be allowed to make. Similarly, you cannot know what moves your opponent can play, because you cannot predict your opponent's dice roll. This kind of game is called **imperfect information**.

III. Game Tree

The turn-based game can be represented as a game tree (Fig. 1). Each node in the tree represents a current state and each branch represents one possible move. Moves lead from one state to another. Each player gets to move at alternating levels of the tree. Because the game is turn based, the state only changes when one player makes a move.

The number of branches from each state is equal to the number of possible moves that the player can make. In many games there can be thousands of possible moves each player can make.

Some states do not have any possible moves. These are called **terminal positions** and they represent the end of the game.

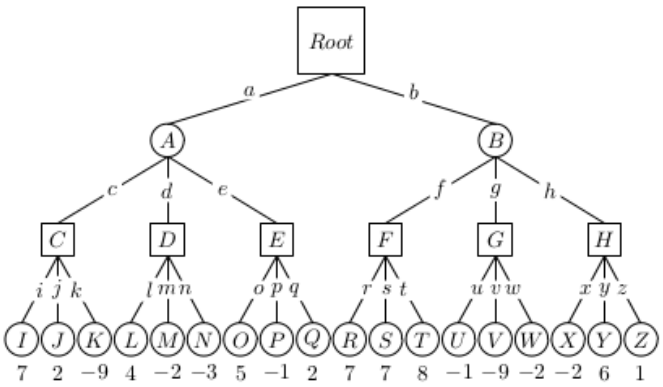


Fig. 1: The game tree [4] of depth of three. Each node in the tree represents a state and each branch move. The leaf nodes represents the states with an computed score using an evaluation function.

Each game has different branch factor and the depth of the tree (maximum turns). Branch factor determines the number of branches on each branch point (node).

It is easier to compute the next move with a small branching factor and deep tree than with a shallow tree but a huge branching factor.

IV. Minmax Algorithm

The Minmax algorithm, sometimes referred as Min-Max or Minimax, is an algorithm introduced by Claude E. Shannon in his seminal paper [4]. When we take a look at Fig. 1 we can imagine root node as an actual state. From this state we can take move *a*, or *b* to make an actual state *A* or *B*. We are on the turn and we are maximizing our score. When our opponent is on the turn, he would like to select the move that maximizes his score too. From our point of view, we don't want to maximize the score, but minimize it. Thus, we can classify the nodes of a game tree as being one of two types: maximizing or minimizing. This observation leads directly to the minmax algorithm. Depending on whether a node is maximizing or minimizing, the algorithm keeps track of the largest or the smallest score, respectively. A leaf node is reached when the remaining depth is equal to zero. At a leaf node, the **evaluation function** is called to determine the score associated with the node.

Minmax algorithm is the most basic algorithm and it is described on the following publications [1–4].

A. Evaluate Function

An evaluate function is function that scores the nodes (states) in the game tree. When the node is in the terminal position the function should return final score. If player is the winner the score is the biggest possible.

In the middle of the game, it is much harder to score. The score should reflect how likely a player is to win the game from the current state. If the board is showing an overwhelming advantage to one player, then that player should receive a score very close to the winning score. In most cases the balance of winning or losing may not be clear [1] .

V. Family of Minmax Algorithms

A. Negamax

The minmax algorithm can be simplified by eliminating the distinction between maximizing and minimizing nodes. By simply negating the result returned from the recursive call, each node can be treated as a maximizing node. At a leaf node, the Evaluate function has to return a score from the viewpoint of the player on move and game should be the zero-sum game.

This modification just simplifies the basic algorithm.

B. Alpha-Beta Pruning

Alpha-Beta pruning (AB pruning) [1], [2], [4] allows the algorithm to ignore sections of the tree that cannot possibly contain the best move. It is made up of two kinds of pruning: alpha and beta.

Together alpha and beta values provide a **window of possible scores**. We (player) will never choose to make moves that score less than alpha, and our opponent will never let us make moves scoring more than beta. The score we finally achieve must lie between the two. As the tree is searched, the alpha and beta values are updated. If a branch of the tree is found which is outside these values, then the branch can be pruned.

This modification significantly speeds up the basic algorithm, although it is heavy dependent on the order of the possible moves.

C. Parallel Alpha-Beta Search

This modification use parallel computation to speed up the AB search and has to deal with three types of overhead: communication overhead, synchronization overhead and search overhead. This type of search introduce several

methods like PVSplit (principal variation splitting), YBWC (young brothers wait concept) and DTS (dynamic tree splitting). For more information read the following article [4].

D. Negascout

Negascout [5] works by doing a full examination of the first move from each state. This is done with a wide search window so that the algorithm doesn't fail. Successive moves are examined using a scout pass with a window based on the score from the first move. If this pass fails, then it is repeated with a full-width window (the same as regular AB pruning). The initial wide-window search from the first move establishes a good approximation for the scout test. This avoids too many failures and takes advantage of the fact that the scout test prunes a large number of branches.

This modification can be faster than AB search, although it is heavy dependent on the order of the possible moves.

E. Aspiration Search

Having a small search window is such a speed up that it can be worthwhile artificially limiting the window. Instead of calling the algorithm with a range of $(-\infty,+\infty)$, it can be called with an estimated range. This range is called an aspiration, and the AB algorithm called in this way is called aspiration search [1]. This smaller range will cause many more branches to be pruned, speeding up the algorithm.

The aspiration for the search is often based on the results of a previous search.

The aspiration search can be used together with AB minimax, AB negamax or Negascout.

F. Memory Enhanced Algorithms

Memory enhanced test algorithms rely on the existence of an efficient transposition table to act as the algorithm's memory. This is simply a zero-width AB search, using a transposition table to avoid duplicate work. The existence of the memory allows the algorithm to jump around the search tree looking at the most promising moves first.

This topic also includes algorithms such as MTD, MTD-f, SSS*, DUAL*, RecSSS*, RecDual, MTD-best...

For more information read the following articles [6][7][1].

VI. Additional optimizations

A. Iterative Deepening

The quality of the gameplay from a search algorithm depends on the number of moves it can look ahead. For games with a large branching factor, it can take a very long time to look a few moves ahead. To avoid being caught without a move, a technique called iterative deepening can be used. Iterative deepening minimax search performs a regular minimax with gradually increasing depths. Initially, the algorithm searches one move ahead, then if it has time it searches two moves ahead, and so on until its time runs out. If time runs out before a search has been complete, it uses the result of the search from the previous depth.

B. History Heuristic

Iterative deepening with memory allows a move to be quickly analyzed at a shallow level and later returned to in more depth. The results of the shallow search can be used to order the moves for the deeper search. This increases the number of prunes that can be made and speeds up the algorithm. Using the results of a previous iteration to order moves is called the history heuristic. It is a heuristic because it relies on the rule that a previous iteration will produce a good estimate as to the best move.

C. Extensions

Extensions are a variable depth technique, where the few most promising move sequences are searched to a much greater depth. By only selecting the most likely moves to consider at each turn, the extension can be many levels deep.

D. Quiescence Pruning

When a period of relative calm occurs, searching deeper often provides no additional information. It may be better to use the computer time to search another area of the tree or to search for extensions on the most promising lines. Pruning the search based on the board's stability is called quiescence pruning.

VII. Conclusion

Negascout combined with Aspiration search and Memory enhanced algorithms are algorithms that can defeat chess world players. Taking into account this fact it seems that there is no room for future improvements.

It is important to note that the most of proposed algorithms deals with speed and efficiency and leaves a room for other areas like the changeable difficulty of the game, alternative moves or even learning from the match. Also methods that are used to speed up the algorithm for example in Memory enhanced algorithms opens the gate for the area of data storage.

We decided to make a benchmark test with selected methods and attributes to better understand the minmax family of algorithms as our next step.

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 1/1147/12 project "Methods for analysis of collaborative processes mediated by information systems".

References

[1] I. Millington, "Board Games", in Artificial Intelligence for Games, T. Cox, Ed. Morgan Kaufmann, 2006, pp. 647-690.

[2] S. Arts, "Competitive play in Stratego," M.S. thesis, Dept. Knowledge Engineering, Maastricht University, Maastricht, The Netherlands, March 2010.

[3] M. Buckland, *Programming Game AI by Example,* Jones & Barlett Learning, 2005, p. 479.

[4] V. Manohararajah, "Parallel Alpha-Beta Search on Shared Memory Multiprocessors," Tech. Rep., 2001.

[5] Tsan-sheng Hsu. (2012). Scout and NegaScout [Online]. Available:http://www.iis.sinica.edu.tw/~tshsu/tcg2012/slides/slide8.pdf

[6] K. Shibahara, N. Inui, and Y. Kotani, "Adaptive strategies of mtdf for actual games." in CIG. IEEE, 2005. [Online]. Available:http://dblp.uni-trier.de/db/conf/cig/cig2005.html#ShibaharaIK05

[7] "SSS* and Dual*." [Online]. Available: http://chessprogramming.wikispaces.com/SSS*+and+Dual*.

Verified Software Development using B-Method

[1]*Zuzana DUDLÁKOVÁ (1st year)*, [2]*Martin VARGA (2st year)*
Supervisor: [3]*Branislav SOBOTA*

[1, 2, 3]Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

[1]zuzana.dudlakova@tuke.sk, [2]martin.varga@tuke.sk, [3]branislav.sobota@tuke.sk

***Abstract*—this paper presents selected parts of a software development using formal method, B-Method. It is divided into two main parts. In the first part, the B-language and construction of verified software is presented. The second part is focused on the design and construction of a railway track using B-Method. Different types of proposals are described. Finally, this article discusses the possibilities of using these proposals and compares different solutions.**

***Keywords*—B-Method, development, formal methods, verification**

I. Introduction

Nowadays, formal methods are used for specification, development and verification of hardware and software. They offer many benefits for software engineering – from code itself to software maintenance. They enable significant reduction of the testing needs and they have excellent potential for automation. [10] Formal methods have the greatest application in the construction of safety critical systems, for example metro and train tracks [9], aircraft systems [4] but they are very useful in the construction of medical heart monitors [3].

This paper presents the use of formal methods, namely B-Method, in safety critical systems. Proposal of the railway track and its possible specification in B language will be presented in this paper.

II. B-Method

The B-Method is a formal method of software development which is based on Z language specification. It was developed by Jean-Raymond Abrial in 1985 – 1988 in Oxford University. It uses a specification and B-AMN (Abstract Machine Notation) as a design language. [1]

The main idea of B is to start with a very abstract model of the system and gradually add details by building a sequence of more concrete models. The goal of B is to obtain a proved model. [2]

B-Method has an industrial tool called Atelier B for the system proposal, from writing, refining and proving specifications to code generation.

A. Life Cycle of Software Development in B

Mathematical proof is an integral part of software development process in the B-Method and that is why all features of the system are still preserved. Abstract formal specification has clearly defined semantics in the initial stages of the development which leads to eliminating errors of system design in the initial state of development. [5]

Stages of the life cycle of software development are presented in Fig.1. [8]

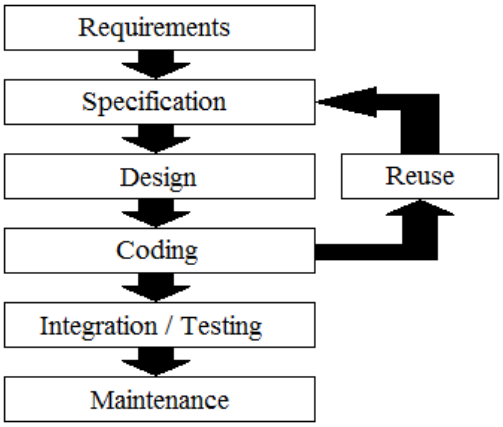


Fig. 1 Life Cycle of Software Development in B

Requirements

This stage consists of creating structured or informal model of soluted problem and requirements for creating system. The result is a set of analytical models.

Specification

Abstract specification in B-Method consists of a number of abstract machines that are mutually interconnected by compositional mechanisms. Each element of a set of analytical models is formalized to form of an abstract machine. Analytical models are used in decomposition of specification to conceptually meaningful components. In next step specification is validated to selected system requirements. Conditions of internal consistency (proof obligations of machine) are generated at the end of the development of specification. They are also proved then. The result is a formal abstract specification of system (MACHINE).

Design

Decomposition of system implementation is identified in the initial stage of design. It includes components from previous development or from specification libraries. The sequence of specification components, which are a type of refinement (REFINEMENT), arises gradual refining of selected

components of a formal specification. The end result is a specification component, which is a type of implementation (IMPLEMENTATION). The proof obligations for created refinements are then generated and proved. It is verification that the refinement refines corresponding more abstract component. The result of this stage is a concrete formal proposal, which consists of one or more specification components (IMPLEMENTATION).

Coding, Integration, Testing

This stage produces an executable implemetation in a specific programming language. The code generator is used for designs at the lowest level (implementation). The generated code is then tested for different model cases based on the system requirements.

B. Language B-AMN

Language B-AMN can be divided into two parts [5]:

a) The language for writing expressions and predicates that define constants, variables, parameters, sets and invariant properties of the specification component. Basis of language is based on classical logic and set theory.

b) The generalized substitution language consists of commands for defining operations in specification component. It describes a set of substitutions that can be used in the language B-AMN.

C. Specification Components in B-Method

The B-Method uses three specification components: abstract machines, refinements and implementation. Individual components are interconnected by compositional mechanizms.

The process of transition from an abstract specification of the system through a series of refinements to the concrete implementation is illustrated in Fig. 2.

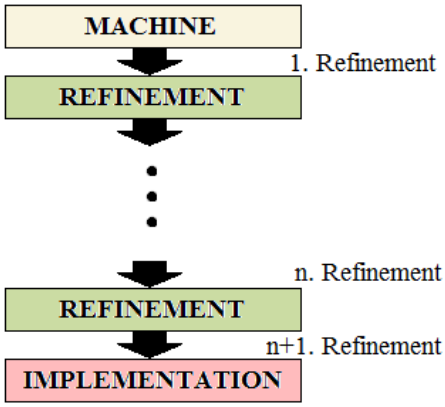


Fig. 2 Sequence of specification components in B

III. Construction of the Railway Track in B-Method

In this section we present the proposal and construction of railway track using B-Method. We used the tool Atelier B.

The study is focused on the construction of railway track with different approaches. The first approach designs the specification without parametrical operations and second approach designs the specification through operations with parameters.

In the Fig. 3 there is a picture which presents proposed model of the track.

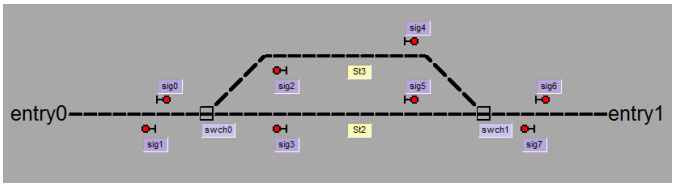


Fig. 3 The Track Layout

The single track consists of two entry points (*entry0, entry1*), which are also represented as signals. It also comprises of two switches (*swch0, swch1*) and eight signals (*sig0, …, sig7*).

The specification is designed to ensure the entry of trains to the track. It also manages the signals and switches and it limits the count of trains in track sections.

A. Specification without parameters

The machine has two types of sets declared: the set SIGNAL (it can take values *green* or *red*) and the set SWITCH (it can take values *switched* or *none*).

The clause VARIABLES consists of variables which represents entry points (*entry0, entry1*), signals (*sig0, sig1, …*) and sections (*entry0_sig1, sig0_swch0, …*). The sections are placed between two different elements in the track. For example: section between two signals (*sig2_sig4*), section between entry point and signal (*entry0_sig1*) and section between switch and signal (*swch0_sig3*).

The clause INVARIANT contains typing of variables. The signals are SIGNAL type, switches are SWITCH type and section are type of 0 or 1 (0 – the section is free, 1 – the section is occupied). It also contains restrictions for variables. They must be valid during lifetime of the system.

The variables are set to initial values in clause INITIALISATION.

Definition of operations occupies the most of the code. In clause OPERATIONS there are specificated operations for every signal (two types of operations for every signal – *requestGreen, requestRed*), operations for entry to the section (they are four for every section, one for entry of train from left side and one for entry of train to the section from right side; the other two are for leaving the section).

The example shows an implementation of operation *reqGreen_sig0* for switching the signal *sig0* to green.

```
  ok<--reqGreen_sig0 =
       PRE
         sig0=red & sig2=red & sig3=red &
         sig1=red & sig7=red & sig0_swch0=0 &
         ((swch0_sig2=0 & sig2_sig4=0) or
(swch0_sig3=0 & sig3_sig5=0))
       THEN
       IF (swch0_sig2=0 & sig2_sig4=0) THEN
           swch0:=switched || sig0:=green
       ELSE /*(swch0_sig3=0 & sig3_sig5=0)*/
           swch0:=none || sig0:=green
           END
           || ok := TRUE
        ELSE
         ok := FALSE
       END;
```

There are restrictions which must be valid before switching the signal to green. In this example there are two possibilities for train because two routes are available. This operation sets the switch to the correct position and sets the signal to green. Signals which are not placed before the switch, the operation only provides a setting of signal to green.

The next example shows an operation which sets an occupation for section *sig0_swch0* when train enters this section.

```
ok<--enterIW_sig0_swch0 =
  PRE sig0=green
  THEN
  IF (swch0=switched) THEN
        swch0_sig2:=1 || sig2_sig4:=1
  ELSE IF (swch0=none) THEN
  swch0_sig3:=1 ||
  sig3_sig5:=1
       END
  END
  || sig0_swch0:=1 || sig0:=red
  || ok := TRUE
END;
```

In this specification there are lots of operations for each element in track. There are defined many conditions for each signal, switch or section which are very similar. That is why we started to design a specification in B with parameters.

B. Specification with parameters

In this part an approach of creating the specification with parameters is presented.

This B machine has more sets: SIGNALS (here are defined all signals – *sig0, sig1, …*), PROP_SIGNAL (this set defines properties of signal, it means the possibility for switch the signal to *green* or *red*), SWITCHES (here are defined all switches placed in track – *swch0, swch1, …*), PROP_SWITCH (it represents the possibility to set the switch to *none* or *switched*), SECTIONS (here are defined all sections – *sig0_swch0, sig2_sig4, …*), PROP_SECTION (this set defines properties of section – *free*, *occup*) and STATIONS (this set contains all stations in track – st2, st3; entries can be consider as stations, too).

In this solution clauses VARIABLES, INVARIANT and INITIALISATION are very similarly defined.

The largest difference is in clause OPERATIONS. This solution is more sophisticated. The operations for setting the signal to red and green have parameters. Operation *reqRed(signal)* has as parameter a signal which has to be switched to red. And operation *reqGreen(signal, station)* has as parameters a signal which has to be switched to green and a final station which says about train routing. It is important to know when track has more then one possible route.

The next example shows an implementation of this solution for operation *reqGreen(signal, station).*

```
ok <-- reqGreen(sem, station) =
  PRE sem:SIGNALS & station:STATIONS
  THEN
   CASE sem OF
EITHER entry0 THEN
 IF(signal(entry0)=red & signal(sig0)=red &
signal(sig1)=red & signal(sig2)=red &
signal(sig3)=red & signal(sig7)=red &
section(entry0_sig1)=free & section(sig0_swch0)=free
& ((section(swch0_sig2)=free &
section(sig2_sig4)=free) or(section(swch0_sig3)=free
& section(sig3_sig5)=free)))
 THEN
  IF(section(swch0_sig2)=free &
section(sig2_sig4)=free)
  THEN
signal:=({entry0, sig0}<<|signal)\/{entry0|->green,
sig0|->green} || switch(swch0):=switched
  ELSE
signal:=({entry0, sig0}<<|signal)\/{entry0|->green,
sig0|->green} || switch(swch0):=none
  END
  || ok:=TRUE
 ELSE ok:=FALSE
 END
OR sig0 THEN
    …
  ELSE ok:=FALSE
  END
END
END;
```

The next operations are defined for smooth transition from station to station. Entries are considered for stations. We have operations: *reqEnter(stationA, stationB)* – it ensures a smooth transition from entry point to the station, first parameter is start station and second parameter is final station; *reqDepartureR(section, station)* – the first parameter is section where train is standing and second parameter is final station, this operation is for trains which are going from left side to right side; *reqDepartureL(section, station)* – this operation is analogical for trains which are going from right side to left side.

We have two other operations for train which are entering to section and for trains which are leaving the section. Operation *enterR(section)* ensures entry of train to section from left side to right side where the section is the input parameter. Analogical operation is *enterL(section)* for trains which are going from right side to left side. The last operation *leave(section)* is responsible for the release section from which the train departed.

IV. Conclusion

It is difficult to say which approach is better. The first approach presents a solution where many operations are needed; they deal with the same thing. For example, the control operation is needed for each signal. This is a great disadvantage. On the other hand, the second approach presents a solution, where a large number of operations are not necessary. For example, two operations (*ReqGreen, ReqRed*) are used for control the signals in the track, where each of them has signal as an input parameter. The disadvantage of this approach is a lot of prooves, that is a big time-consuming.

These proposed specifications are useable with tool Train Director [6], [7].

The B-Method is one of the most recognizable formal methods. Therefore, we just started with it. Other very popular methods are Event B[11] and Perfect[12]. In the future we plan to use them in our system proposal.

Acknowledgment

This work has been supported by KEGA grant project No.

050TUKE-4/2012:"Application of Virtual Reality Technologies in Teaching Formal Methods".

REFERENCES

[1] J. R. Abrial, The B Book: Assigning programs to Meaning. New York: Cambridge University Press, 1996, ISBN 0-521-49619-5.
[2] D. Bjorner, M. C. Henson, Logics of Specification Languages, Springer Berlin Heidelberg, 2007.
[3] V. Gehlot, E. B. Sloane, Software and System Engineering to Ensure Patient Safety in Wireless Medical Device Networks, Villanova University, Philadelphia, Pennsylvania, 2006.
[4] A. Joshi, S. P. Miller, Mode Confusion Analysis of a Flight Guidance System Using Formal Methods, Proceedings of the 22st Digital Avionics Systems Conference, Indianapolis, Indiana, October 12-16, 2003, Department of Computer Science and Engineering, University of Minnesota, Minneapolis, 2003.
[5] Š. Korečko, O. Látka, C. Szabó, F. Hrozek, Riešenie úloh spracovania rozsiahlych grafických údajov v prostredí paralelných počítačových systémov, KPI FEI TU Košice, 2012, ISBN 978-80-553-0864-7.
[6] Š. Korečko, B. Sobota, C. Szabó, Using Simulation and 3D Graphics Software to Visualize Formally Developed Control Systems, Proceedings of the 15th WSEAS International Conference on Computers, Corfu, Greece, July 14-17, 2011, Piraeus, Greece, WSEAS Press, 2011.
[7] Š. Korečko, J. Sorád, B. Sobota, An External Control for Railway Traffic Simulation, Proceedings of the Second International Conference on Computer Modeling and Simulation, Brno, Czech Republic, September 5-7, 2011, Faculty of Information Technology, Brno University of Technology, 2011.
[8] K. Lano, The B Language and Method: A Guide to Practical Formal Development, Springer-Verlag London, 1996.
[9] T. Lecomte, T. Servat, G. Pouzancre, Formal Methods in Safety-Critical Railway Systems, Clersy, Aix en Provence, France, 2010.
[10] C. Popescu, J. L. Martinez Lastra, Formal Methods in Factory Automation, InTechOpen, March 1, 2010, Tampere University of Technology, Finland, 2010.
[11] Event B homepage: http://www.event-b.org/
[12] Perfect homepage: http://www.eschertech.com/

Section:
Electrical & Electronics Engineering
Oral form

An OPNET Modeler based simulation approach for wireless video transmission

[1]*Lukáš SENDREI (1^{st} year)*
Supervisor: [2]*Stanislav MARCHEVSKÝ*

[1,2]Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

[1]lukas.sendrei@tuke.sk, [2]stanislav.marchevsky@tuke.sk

***Abstract*—One of the biggest challenges of communication networks is the video transmission in real time. It requires high demands on the available network capacity and transport mechanisms. Availability of smart mobile devices with batteries, which keep the terminal working for several hours, caused an increased interest in the research of the deployment of video transmission in wireless transmission systems. The presented paper deals with the transmission of video encoded with H.264/AVC (Advanced Video Coding) video coding standard through wireless local area network (WLAN) using the programming environment OPNET Modeler (OM). The test network studied in this work was prepared by combining real and simulated networks, which allows interesting possibilities when working with the OM tools. Such an approach to working with OM allows a detailed video streaming analysis, because the video output is noticeably not only in the form of statistics, but one can see the real impact of transmission failures. The OM simulation environment allows design transmission systems, which would be difficult to establish in laboratory conditions.**

***Keywords*—H.264/AVC, OPNET Modeler, system-in-the-loop, wireless local area network**

I. Introduction

The development of wireless communication has been marked by a huge expansion in recent years. We deal with it every day. It simplifies our work, communication and entertainment. The achievements of modern mobile networks reach parameters which ensure quality delivery of information to various applications. They found wide usage in many areas. This creates significant pressure on the efficient use of available resources according to the philosophy "transfer as much as possible for as little as possible."

Despite the advanced communication technology the research is also focused on the video coding compression algorithms. In this area, several standards for video coding were created. A number of research institutions, organizations, universities and also commercial companies participate in the time-consuming and difficult process of standardization. One of the latest standards for video coding is H.264/AVC, which represents a significant movement in terms of reducing coding requirements and improves visual quality [1], [2].

This paper shows the advantageous combination of a real network with a simulated WLAN network designed in OPNET Modeler environment. The real video transmission generated by the video server was imported into the simulated network, transmitted through the simulated WLAN network and transported back to the real network, where the received video sequence was decoded. OPNET Modeler was implemented as a generator of statistical parameters.

This paper is organized as follows. The second part is dedicated to a brief review of the H.264/AVC video compression standard. In the third part the OPNET Modeler is introduced and the possibilities of video transmission in this environment are described. Part four deals with the simulations and summarizes the results. The last section contains a brief summary.

II. H.264/AVC

H.264/AVC is an international video coding standard that was first published in 2003. The standard was created within the collaboration of the working group (WG) MPEG (Moving Picture Experts Group) of organization ISO/IEC and VCEG group (Video Coding Experts Group) of the ITU-T institute. This WG's created together a WG JVT (Joint Video Team) to finalize the project of ITU-T called H.264 and ISO/IEC project named MPEG-4 part 10, Advanced Video Coding. H.264/AVC represents the latest concept of video coding from the series of standards H.261, MPEG-1 Video, MPEG-2 Video, H.263 and MPEG-4 Visual/Part 2. The main aim of the standardization effort was to achieve increased coding efficiency and provide a better adaptation to the transmission network for various applications like videoconference, video streaming, and video on demand, TV broadcasting or video storage. Compared with older standards, H.264/AVC provides a more efficient compression of different bit rates and video resolution by 50% on average. But this resulted into more complex decoder architecture [1], [3], [4].

The standard defines the syntax and semantics of the bit stream as well as pre-processing which requires the decoder to correctly decode the bit stream to a viewable video. The standard does not define the H.264/AVC video coding or other pre-processing processes, enabling manufacturers to operate at a price, coding efficiency, error concealment, error resiliency or hardware requirements. Another area where different manufacturers can use their tools is the processing of the decoded video to create a video stream optimized for the target application [2]. The concept of H.264/AVC is based on two concept layers [1], [5]:

- **VCL** (Video Coding Layer) – executes the video coding itself
- **NAL** (Network Abstraction Layer) – supports video transmission trough different types of networks

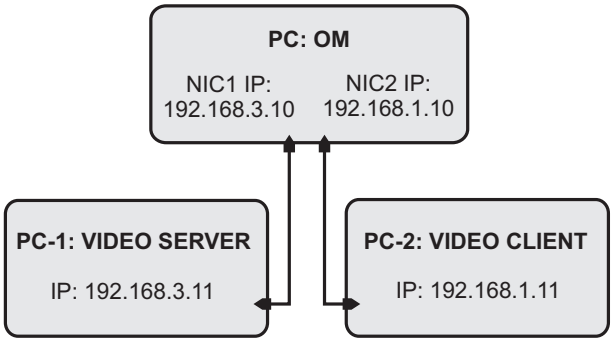


Fig. 1. The topology of the test network with three workstations.

The design of VCL is based on hybrid block video coding. The basic coding algorithm is a combination of inter frame prediction to exploit temporal statistical dependence and transform coding of pre-prediction residuals to exploit spatial statistical dependencies. In the VCL there is no single coding element present that can provide the most significant improvement in compression efficiency than in previous standards. It is mainly a number of small improvements that together produce significant efficiency gain of the video coding algorithm. NAL layer is designed to organize video VCL layer data and provides header information that is used for suitable video transmission over the transport layer or storage. It facilitates the ability of mapping the VCL data to transport layers of many transmission systems [1].

III. OPNET Modeler

OPNET Modeler is one of the most powerful simulation tools in the field of communication technologies. It is deployed in research and development of communication protocols and devices. Designers and administrators of communication networks can use this tool for efficiency analysis, optimization, and analysis of future growth and other properties of networks. Sophisticated network analysis suite of tools available for OPNET Modeler provides the opportunity to specify the models in details, identify significant individual components, perform simulations and analyze the generated output data [6].

System-in-the-loop (SITL) is a special module that allows connection of a simulated network in the OM environment and software applications or hardware of networks operating in real time. By using SITL during the simulation, packets are transmitted between the simulated and real network in real time. SITL Gateway in OM represents an external device which exchanges simulation packets and the WinPcap library, in the operating system Windows, is used for routing user-defined filter selected packets from the Ethernet network adapter in the simulation. The workstation on which OM runs, and hence the simulation, is connected to the real network via Ethernet interface. The simulation must run in real time to guarantee the synchronization and conversion of packet traffic flows inside the simulated system.

There are three strategies for simulations with SITL [7]:

- **real-to-sim** - communication between real and simulated network
- **real-sim-real** - communication between real networks through simulated network
- **sim-real-sim** - communication between simulated networks through real network

OM contains a range of predefined applications that can be adjusted when necessary. Video transmission is provided by the application named videoconferencing. We can set a frame size and frame rate and even some network parameters. This represents an opportunity to compile the required data flow model of video sequences, but does not allow visualization and change of the specific parameters of the video streams. An interesting solution in this respect is the connection of real systems with the simulation in the OM environment. Using this variant we can connect real network device that will generate and also receive real video streams to simulated network that will transport these streams during the simulation. According to the selected characteristics, type and topology the simulated network affects the transmission parameters of the real video content. Thus it is possible in the simulation environment to design very complex and extensive networks and it is easy to test them with real applications.

IV. Simulations

Based on the facts mentioned above, an approach that considers the connection between the real equipment and a simulated network in the OM environment has been chosen. SILT ensures the connection between the real workstations and the simulation. OM provided video transmission through the simulated WLAN. The test topology consisted of three workstations. The video server was on the first workstation *PC-1: VIDEO SERVER*. Workstation *PC-2: VIDEO CLIENT* displayed and also stored the transmitted video using VLC media player version 1.1.11. Workstation *PC: OM*, where OM was implemented, has two network interfaces, to which the workstation server and client were connected. The topology of a test network is shown in Fig. 1.

The simulation was carried out in the programming environment OM. The simulated WLAN network was created in the OM project editor. The WLAN was composed of two wireless routers. The wireless router *WLAN AP* was the Access Point (AP) and the AP functionality was disabled on the second router *WLAN Receiver*. Both WLAN routers were connected using the special link *sitl_virtual_eth_link* with the proprietary interface *sitl_virtual_gateway_to_real_world*. They both are a part of the SITL module object library. Fig. 2 shows the simulated WLAN network design in OM.

Three different video resolutions were used, 1920x1280, 1280x720 and 640x360 pixels. The videos were converted using the *x264* codec. All videos were saved in mp4 format. The video was 3554 frames long and the frame rate was 24 fps (frames per second). The video server created packets that were transported by UDP (User Datagram Protocol) in the

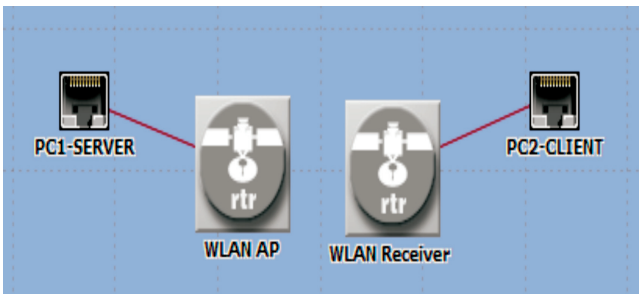


Fig. 2. Simulated Wireless LAN in OPNET Modeler.

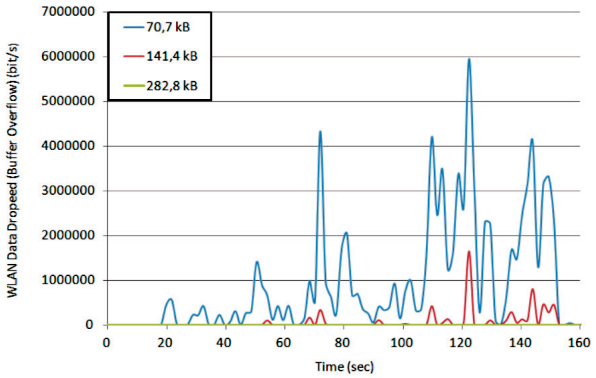


Fig. 3. WLAN buffer overflow (Video resolution 1920x1080 pixels).

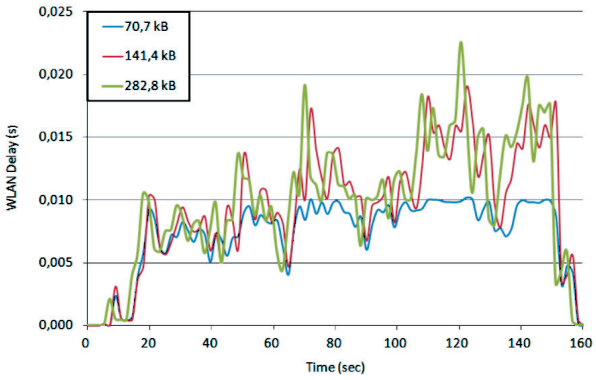


Fig. 4. WLAN delay (Video resolution 1920x1080 pixels).

form of variable bit rate (VBR). The maximum size of the packets was 1414 B.

During the simulation the WLAN IEEE 802.11g technology [8] was used at the transmission speed of 54 Mb/s, with the exception of the video transmission testing at lower WLAN data rate.

The video transmission by the resolution of 1920x1080 pixels was compared for three WLAN router buffer size: 70.7 kB, 141.4 kB and 282.8 kB. The WLAN data dropped amount for three different buffer sizes for the transmission of HD video (1920x1080) is shown in Fig. 3. Sufficient memory size for lossless HD video transmission through WLAN at a resolution of 1920x1080 pixels is 282.8 KB. For this size and resolution we can achieve lossless transmission. The delay in the transmission of HD video 1920x1080 is visible in Fig. 4. The WLAN data dropped amount for HD video resolution 1920x1080 at low data rates is shown in Fig. 5. The results of the simulation show that the necessary transmission speed of wireless network is at least 12 Mb/s.

The buffer size by the video transmission at lower resolution, namely 1280x720 pixels, was: 35.35 kB, 70.7 kB and 141.4 kB. Comparison of WLAN data dropped amount of the video stream is visible in Fig. 6 and the transmission delay in Fig. 7. This simulation shows that for lossless transmission of video of that resolution (1280x720) it is necessary to have a buffer size of 141.4 kB. Comparison of transmission losses at WLAN data rates 2, 6 and 9 Mb/s contains Fig. 8. The lowest bit rate, where significant losses were recorded was at 6 Mb/s. The lowest bit rate, which causes no losses by video streaming through WLAN of this resolution, is 9 Mb/s.

The properties of the transmission of low video resolution 640x360 pixels were tested for AP buffer size of: 7.07 kB, 21.21 kB and 42.42 kB. Lossless transmission allows the size of 42.42 kB. Comparison of losses by the video transmission of resolution 640x360 pixels with varying buffer size is summarized in Fig. 9 and the relayed delay in Fig. 10. Relatively large losses and the associated delays are still reported at a rate of 1 Mb/s. Almost lossless transmission is possible by transmission speed of 2 Mb/s and completely lossless transmission at 5.5 Mb/s. Results of the loss comparison at WLAN data rate 1, 2 and 5.5 Mb/s for low resolution 640x360 pixels is shown in Fig. 11.

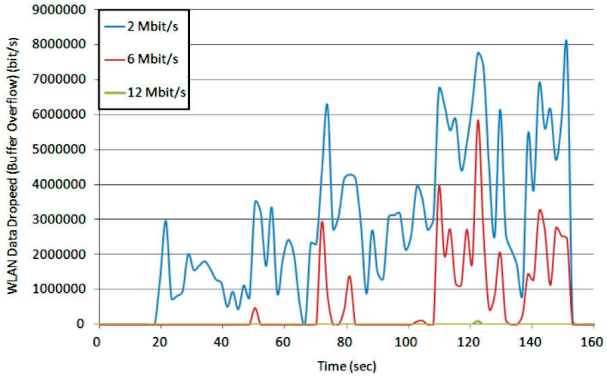


Fig. 5. Losses by low WLAN data rates (Video resolution 1920x1080 pixels).

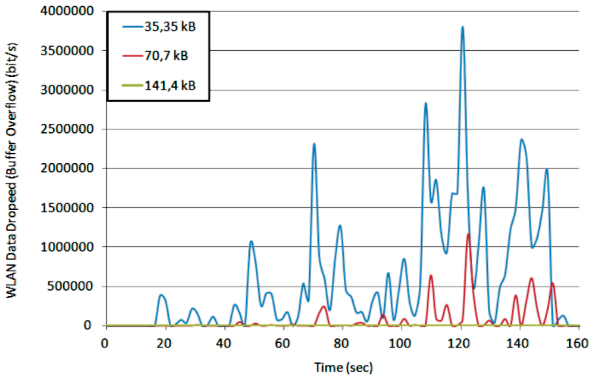


Fig. 6. WLAN buffer overflow (Video resolution 1280x7200 pixels).

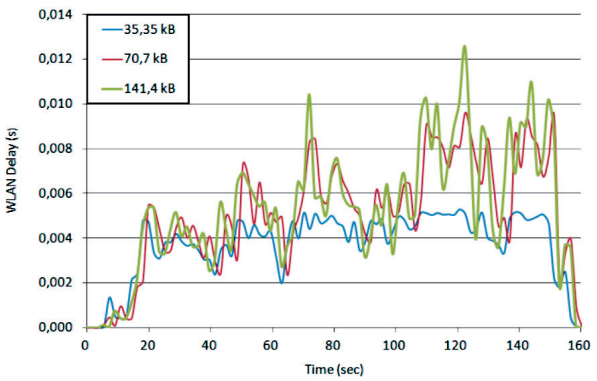


Fig. 7. WLAN delay (Video resolution 1280x720 pixels).

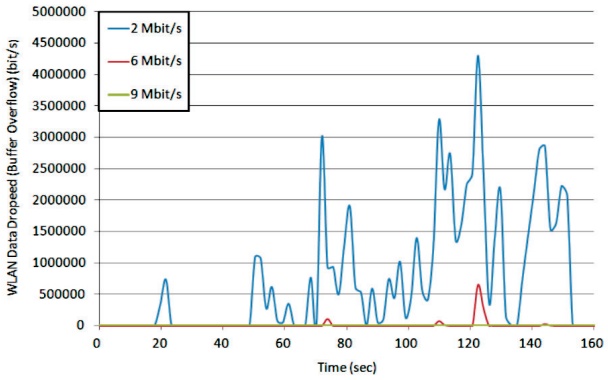


Fig. 8. Losses by low WLAN data rates (Video resolution 1280x720 pixels).

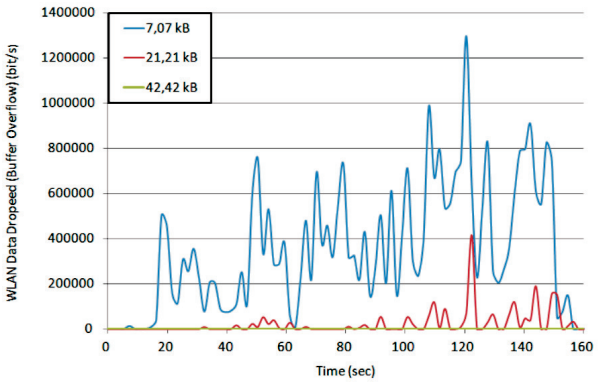


Fig. 9. WLAN buffer overflow (Video resolution 640x3600 pixels).

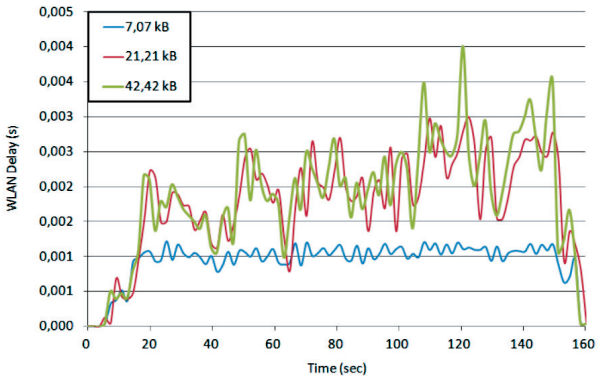


Fig. 10. WLAN delay (Video resolution 640x360 pixels).

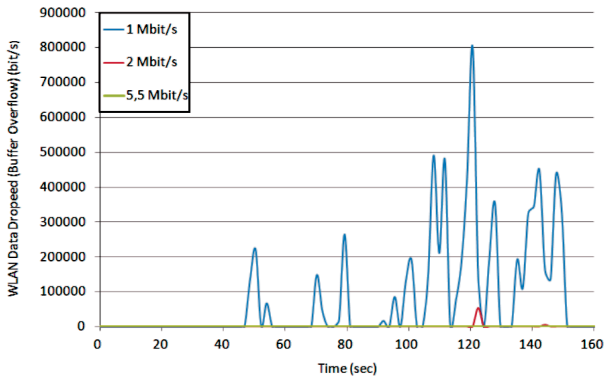


Fig. 11. Losses by low WLAN data rates (Video resolution 640x360 pixels).

V. Conclusion

The amount of simulated WLAN load scenarios with different parameter settings showed the possibilities and limitations in the video transmission of various resolutions in the OM environment. The H.264/AVC standard has been proven to be a suitable video encoding format to ensure the highest possible video stream transmission quality with relatively low requirements on the communication network. The shortcomings of the transmission of VBR video stream can be removed either in the process of video encoding, or by ensuring fixed bit rate on the video server site. That solution would reduce the load of network components. The delay variations can be sufficient and in balance using cache memory on the video decoder side. The only problem at the current state of technology, which cannot be satisfactorily solved, is the transfer of HD video at low bit rates. Further information about the topic of video transmission over WLAN's and extensieve simulations can be found in [9].

Acknowledgment

This work is the result of the project implementation: Development of the Center of Information and Communication Technologies for Knowledge Systems (ITMS project code: 26220120030) supported by the Research & Development Operational Program funded by the ERDF.

References

[1] T. Wiegand, G. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 13, no. 7, pp. 560–576, July 2003.

[2] J. Ostermann, J. Bormans, P. List, D. Marpe, M. Narroschke, F. Pereira, T. Stockhammer, and T. Wedi, "Video coding with H.264/AVC: tools, performance, and complexity," *Circuits and Systems Magazine, IEEE*, vol. 4, no. 1, pp. 7–28, 2004.

[3] I. Richardson, *The H.264 Advanced Video Compression Standard*. John Wiley & Sons, 2010.

[4] T. Wiegand, H. Schwarz, A. Joch, F. Kossentini, and G. Sullivan, "Rate-constrained coder control and comparison of video coding standards," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 13, no. 7, pp. 688–703, July 2003.

[5] G. M. Ghandi M., "The H.264/AVC video coding standard for the next generation multimedia communications," *J. Iranian Association Elec. Eng.*, vol. 1, no. 2, pp. 3–14, 2004.

[6] J. Mohorko, F. Matjaz, and K. Sasa, "Advanced modelling and simulation methods for communication networks," *Microwave Review*, vol. 14, pp. 41–46, September 2008.

[7] O. Technologies Inc., "OPNET Modeler product documentation release," 2008.

[8] M. Etoh, *Next Generation Mobile Systems: 3G & Beyond*. John Wiley & Sons, 2005.

[9] L. Sendrei, "H.264 video transmission in wireless local area networks using Opnet Modeller," Master's thesis, Technical University of Kosice, Faculty of Electrical Engineering and Informatics, Department of Electronics and Multimedia Telecommunications, 2012.

Analogue pulse generator for Multiphase Boost Converter

[1]*Ján PERDUĽAK,* [2]*Victoriia KOVALCHUK*
Supervisor: [3]*Dobroslav Kováč*

[1,3] Dept. of Theoretical Electrical Engineering and Electrical Measurement, FEI TU of Košice, Slovak Republic
[2] Institute of Electromechanics, Energy Saving and Control Systems Kremenchuk Mykhailo Ostrohradskyi National University, Ukraine

[1]jan.perdulak@tuke.sk, [2]viktoria_kovalc@mail.ru

***Abstract*— this article deals with design control for novel concept of multiphase boost converter. This new concept allows effective utilization of energy from input source. The effective utilization of input energy is ensured by adding five parallel legs to the conventional single phase boost converter. Appropriate algorithm of switches control allows converter to take the input source energy one of these six parallel legs in every moment. The suitable algorithm of switches control ensures that almost whole input source energy is effective utilized. The simulation and experimental models have been built and measured to verify the theoretical properties of design control and correct function of multiphase boost converter.**

***Keywords*—single phase boost converter, multiphase boost converter, logical operation, exclusive OR.**

I. INTRODUCTION

This paper presents the novel concept of control of multiphase boost converter *(MPBC)* with high efficiency of energy conversion. The high efficiency of energy conversion is ensured by adding five more parallel legs to the conventional single phase boost converter (*SPBC)*. The suitable algorithm of switches control in particular legs ensures that the almost whole input source energy to the converter is effective utilized.

The principle of function of conventional *SPBC* is well known. When the switch *S* is turned on the input energy starts to accumulate in form of magnetic field in the inductor *L*. This accumulated energy included with source energy is delivered to the output *Z* after switch *S* is turned off. The average value of the output voltage $U_{Z(AV)}$ in continuous conduction mode *(CCM)* is:

$$U_{Z(AV)} = \frac{1}{1-z} U_{IN}. \qquad (1)$$

where duty cycle "*z"* is ratio between time when the switch *S* is turned on $t_{on(S)}$ and the period *T*, $z = t_{on(s)}/T$. The theoretical waveforms and topology of conventional *SPBC* are shown in fig.1.

We can notice that there exists a time interval within the period *T* when the energy delivered to the load *Z* is equal zero. This is the main problem of conventional *SPBC* - effective utilization of input energy. We have to ensure that the input energy will be delivered to the load Z over the whole period *T*. We have to remove the time interval within the period *T* where the energy delivered to the load *Z* is zero.

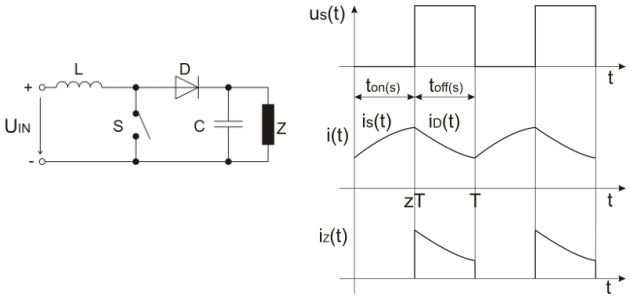


Fig. 1. Conventional topology and theoretical waveforms of SPBC.

This is done by using the proposed *MPBC*, fig.2. The abovementioned process of delivering energy to the load *Z* can be repeated 6-times because 6- parallel legs are presented in *MPBC*. The proposed *MPBC* allows elimination the time intervals when the energy delivered to the load *Z* is zero and thus effective utilization of energy from input source.

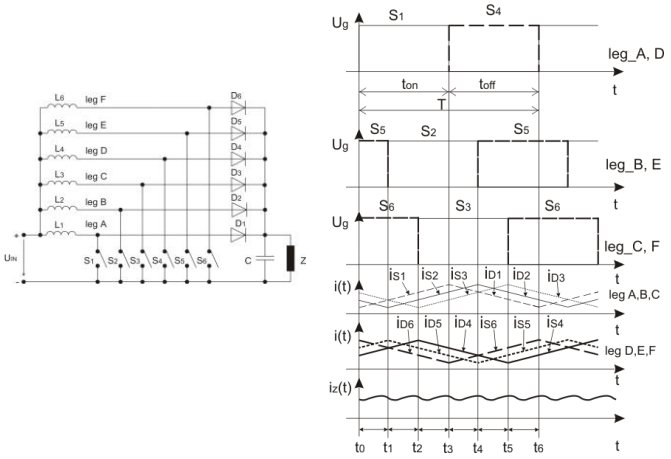


Fig. 2. Topology and theoretical waveforms of MPBC.

II. Design Control

The correct function of *MPBC* is ensured by suitable algorithm of design control which allows effective utilization of converter input energy and eliminate the time interval when the converter output energy is zero.

The main idea of design control is to create the six phase shifted pulses to cover the whole period *T* of delivering the converter input energy to the load *Z*.

The flowchart and proposed control structure are shown in fig.3 or more precisely in fig.4.

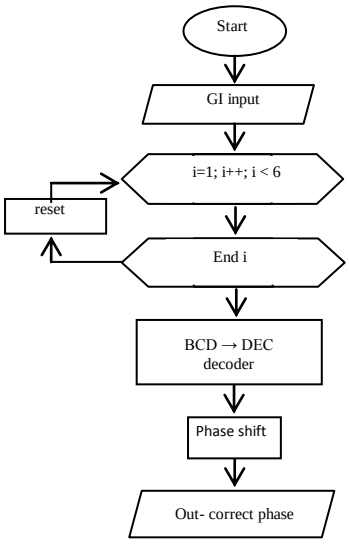


Fig. 3. Block diagram of design control.

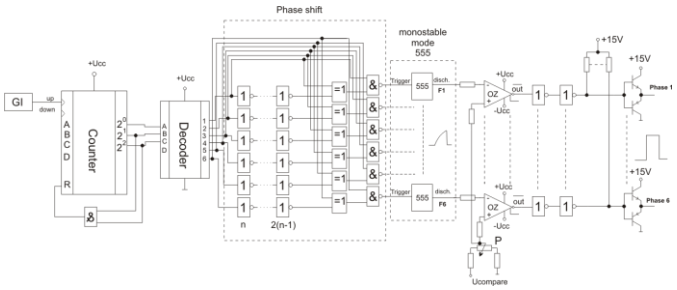


Fig.4. Proposed control structure of MPBC.

Pulse generator signal is connected to the input of counter. The counting of counter is limited to six because six phase of MPBC. The operating frequency of MPBC is set to f_S = *50 kHz* so the frequency of pulse generator signal has to be set at f_{PG} = *300 kHz*.

The counter output signal enters to the *BCD* (**B**inary **C**oded **D**ecimal) to decimal decoder. The decoder ensures six output signals to control six phases of *MPBC*. The correct phase shift is ensured by means of basic logical operation executing by basic logic elements in block Phase Shift. The phase shifted signals enter to the 555 timer which works in monostable mode. Triangle signal is taken from discharge pin of 555 timer and enters to the operational amplifier which works as comparator. Operational amplifier *(OA)* compares the triangle signal with voltage $U_{compare}$. The voltage level $U_{compare}$ is sets by means of potentiometer *P* and thus the different values of duty cycle *z* can be set. At the end of this process the *OA* outputs are set to desired level to control the switches in particular legs of *MPBC*.

III. Simulation Results

The simulation model of proposed control structure of *MPBC* converter shown in Fig.5 was created in simulation environment OrCAD Capture CSI to verify its theoretical properties.

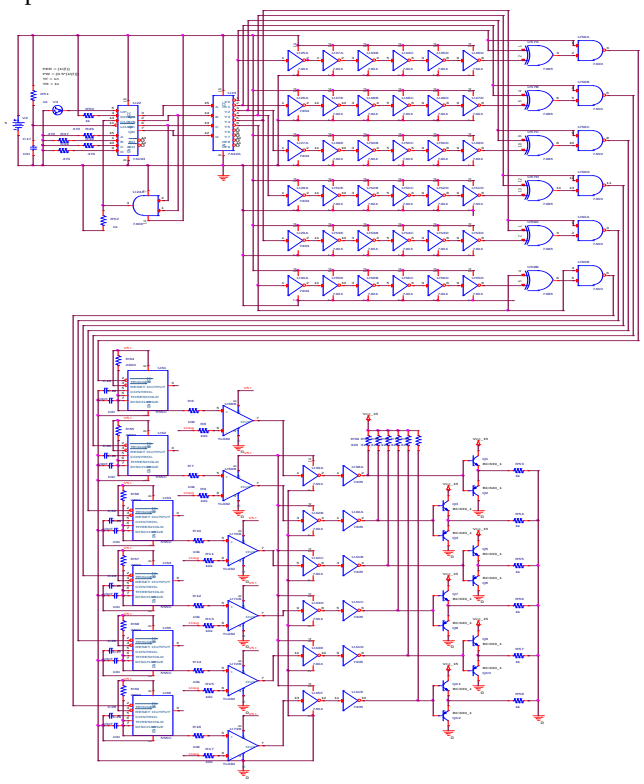


Fig.5. Simulation model proposed control structure.

Pulse generator output enters to the simulation model (input *UP*) of counter *IO 74 193*. The counting of counter is limited to six because of six phases of converter. Outputs (Q1 – Q3) of *IO 74 193* enter to the inputs of simulation model of DCB to decimal decoder *IO 7442*. Fig. 6, shows the pulse generator signal (PG), output signals from counter *IO 74 193* (Q1 – Q3) and output signals from decoder *IO 7442* (Y0 – Y5).

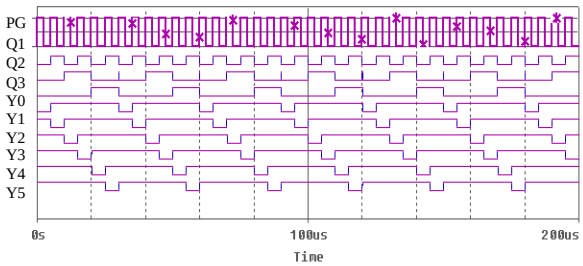


Fig.6. Output signals form pulse generator (PG), *IO 74193* (Q1 – Q3) and *IO 7442* (Y0 – Y5).

The correct phase shift is ensured by means of basic logical operation executing by basic logic elements in block Phase Shift.

Fig.7 shows the logical operation exclusive *OR (XOR)* of output signal *Y0* with shifted output signal *Y0_shift*. The results of logical operation *XOR* are two short pulse signals, depicted in Fig.7.

Fig.8 shows the logical operation *NAND* of output signal *Y0* with two short pulse signals as a result of previous logical operation *XOR*. The result of logical operation *NAND* is

inverted signal which serves as an input trigger signal to 555 timer that works in monostable mode.

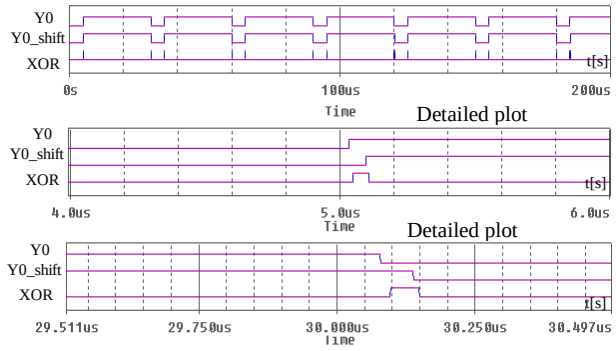


Fig.7. Ligical operation exclusive OR.

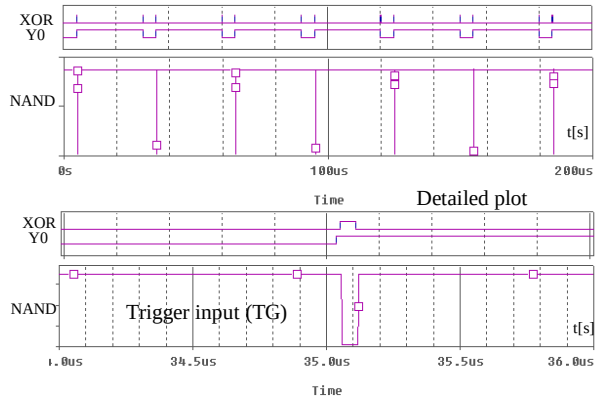


Fig.8. Logical operation NAND.

TABLE I
TRUTH TABLE OF XOR AND NAND

A	B	A XOR B	A NAND B
0	0	0	1
0	1	1	1
1	0	1	1
1	1	0	0

Fig.9 shows all six phase shifted signals entering to the trigger input *(TGI)* of timer 555 and the output signals are taken form discharge *(DISCH)* pin of timer 555. It can be seen that all discharge signals are triangular.

Triangle signal is taken from discharge pin of 555 timer and enters to the *OA* which works as comparator. *OA* compares the triangle signal with voltage $U_{compare}$ and the outputs of operation amplifier are set to desired level to control the switches in particular legs of *MPBC*. Fig.10 (lower part) shows all six control voltage for particular switches of *MPBC*.

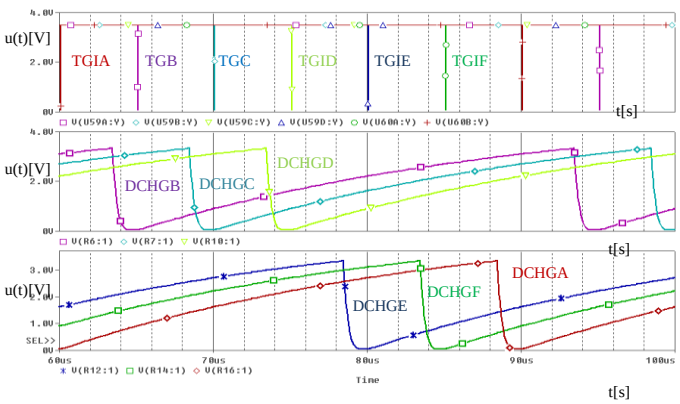


Fig.9. Trigger inputs to 555 timer (upper part) and discharge outputs (lower part).

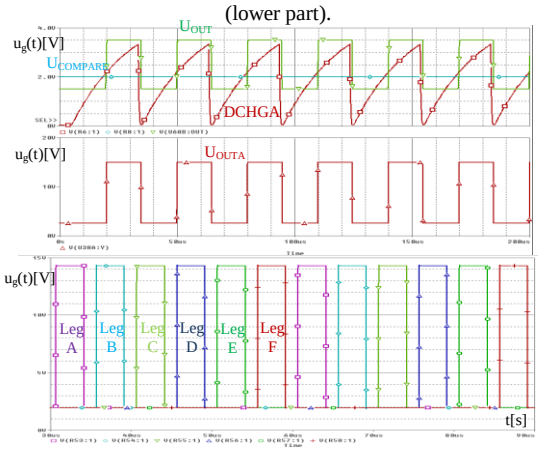


Fig.10. comparative process (upper part), comparator output (middle part) and phase shifted control voltage for particular switches of MPBC (lower part).

IV. EXPERIMENTAL RESULTS

The laboratory model of control structure has been built and tested to verify the principle of the operation.

The following oscillograms fully confirm the theoretical and simulation assumes and thus only the particular oscillograms and its description are placed in this part.

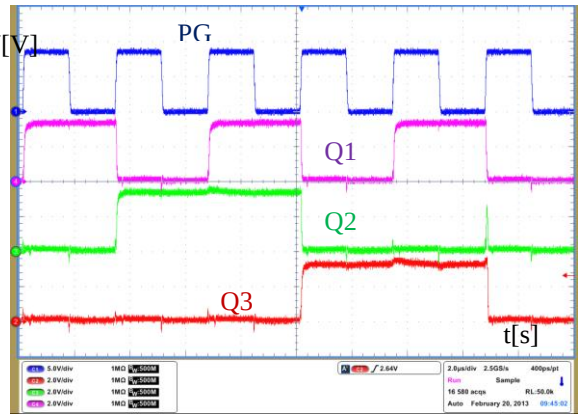


Fig.11. Output signals from PG and counter *IO 74193* (Q1 – Q3).

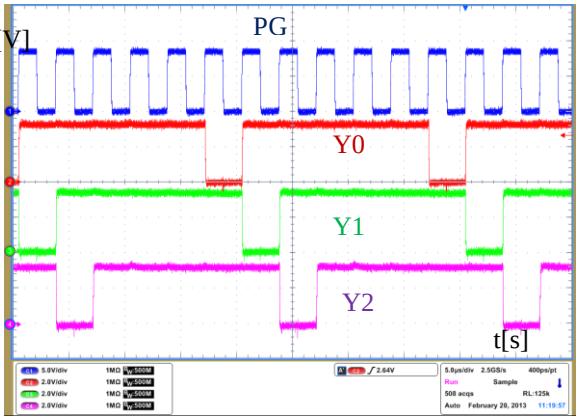


Fig.12. Output signals from PG and decoder *IO 7442* (Y0 – Y2).

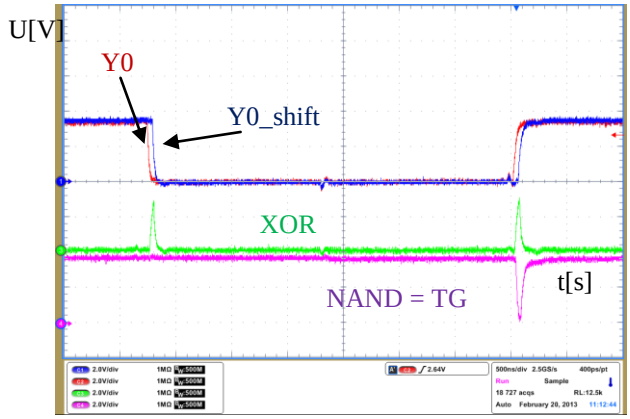


Fig.13. Logical operation XOR and NAND.

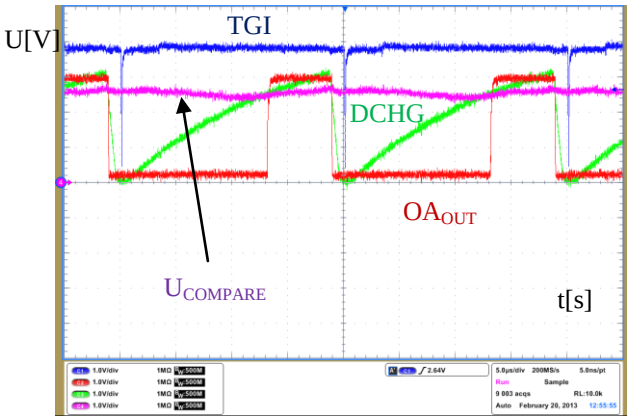


Fig.14. Trigger inputs to 555 timer and discharge outputs.

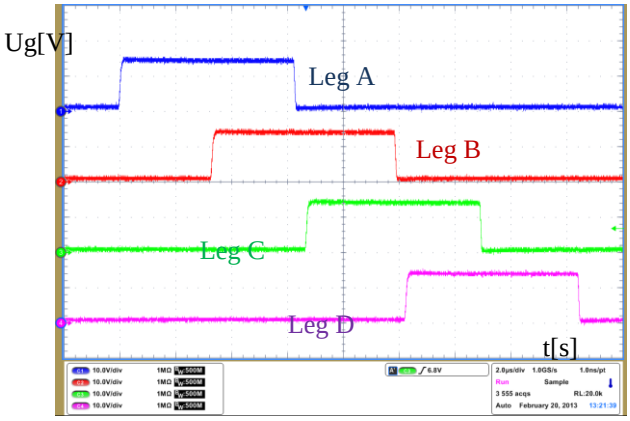


Fig.15. Phase shifted control voltage for particular switches of MPBC.

The designed analogue generator has been interconnected with *MPBC* and experimental results have been obtained as shows Fig.16. The input voltage U_{IN} was set to 6V. The duty cycle *z* was set to 60 percent of period *T*. It can be seen that the almost whole delivered input energy is utilized by this converter. This is due to switches in particular legs. The converter operates in *CCM* and continually delivers input source energy to the load *Z* in every instant of time.

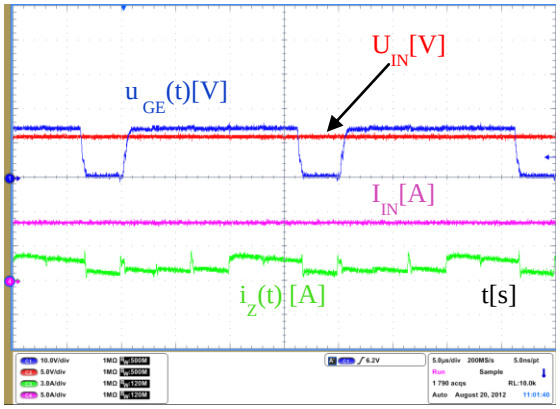


Fig.16. Waveforms of gate voltage $u_{GE}(t)$,load current $i_Z(t)$, photovoltaic voltage and current U_{PV} and $i_{PV}(t)$ at U_{PV}=6V

V. Conclusion

The *MPBC* works with high efficiency of energy conversion in compare with *SPBC*. The *MPBC* continually delivers the input source energy to the load *Z* using six parallel phases. In this case, the time interval when the output energy is equal to zero is removed.

Acknowledgment

The paper has been prepared under support of Slovak grant projects KEGA No. 005TUKE-4/2012 and KEGA No. 014TUKE-4/2013.

References

[1] D. Kováč, I. Kováčová, Patent application No. 00150-2010 titled as, *Multiphase boost DC/DC converter.*

[2] D. Kováč, I. Kováčová, Patent application No. 00001-2011 titled as, *Multiphase boost DC/DC converter with constant slope falling – off of inductance current.*

[3] D. Kováč, I. Kováčová, J. Perduľak, T. Vince, J. Molnár, Patent application No. 00097-2011 titled as, *Pulse generator for multiphase boost converter.*

[4] D. Kováč, I. Kováčová, J. Perduľak, T. Vince, J. Molnár, Patent application No. 00097-2011 titled as, *Analogue pulse generator for multiphase step-up DC/DC converter*

Development of Sensoric Subsystem for Physical Model of Helicopter

[1]*Ján BAČÍK (1st year)*, [2]*Radovan SIVÝ (1st year)*
Supervisor: [3]*Pavol FEDOR*

[1,2,3] Department of Electrical Engineering and Mechatronics, Technical University of Košice, Slovak Republic

[1]jan.bacik.2@tuke.sk, [2]radovan.sivy@tuke.sk, [3]pavol.fedor@tuke.sk

***Abstract*— The paper deals with hardware design of sensoric subsystem for physical model of helicopter that is characterized by a long-term stability and in real time it generates data about accelerations, angular rates and position during helicopter fly. In our paper we used small model helicopter T-REX 600 for testing sensoric subsystem in varying conditions. The sensoric system is based on powerful 32-bit processors with the cores ARM7 and Cortex-M3. The main unit for data processing presents an embedded computer built on a mini-ITX motherboard with the processor Intel i3. The helicopter presents a system with six degrees of freedom. In the fact, during the flight, there is not any fixed point that would enable to calibre the sensors placed on the helicopter board, so for processing sensor data complex stochastic calculations are necessary. They are based on a discrete Kalman filter.**

***Keywords*— Kalman filters, sensoric system, helicopter**

I. Introduction

Navigation presents a very old art which has become a complex science. It is essentially about finding the right way of a body from one place to another and there exist many possibilities to achieve this objective.

The operation of inertial navigation systems depends upon laws of classical mechanics as were formulated by Sir Isaac Newton. For example, if given the ability to measure acceleration, it would be possible to calculate the change in velocity and position by performing successive mathematical integration of the acceleration with respect to time. In many sensoric systems the inertial sensors are mounted on a stable platform and are mechanically isolated from the rotational motion of the vehicle. Modern sensoric systems have removed most of the mechanical complexity of platform systems by attaching the sensors rigidly, or "strapped down" to the body of the vehicle. The potential benefits of this approach are lower cost, reduced size and greater reliability. The major disadvantage consists in increase of computing complexity. At present the tasks of strapdown inertial navigation [1] are used in modern robotics systems.

Current development of considerably cheap MEMS sensors [2] and powerful processors enable to implement techniques of inertial navigation [3, 4] into a light-weight and powerful equipment. One of new promising way is based on utilization of small commercial mini-ITX boards with the Intel processors that allow to install an arbitrary operational system and to utilize its advantages for data processing.

In the contribution we describe development of a sensoric system hardware placed on board of a real helicopter model (Fig. 1) enabling to process and evaluate flight data. Based on input vector and sensed output vector components it is possible to derive a fuzzy model of the system. Computing the state vector based on data collected from various sensors is performed in real-time. This solution was chosen due to a possible implementation of the sensoric system into a real-time control system what we are going to perform in next step of our research.

Fig. 1 Small model of a helicopter T-REX 600

The debugged program is based on the Linux operational system, utilization of the GNU Scientific Library [5], and subsystems utilizing powerful 32-bit processors with the cores ARM7 and Cortex-M3 [6].

II. Equations for Strapdown Inertial Navigation of the Body in Space with 6 Degrees of Freedom

As it is mentioned above the inertial navigation is based on Newton differential equations that can be expresses in the form (Eq. 1-4):

$$\dot{P}_e = T_{BE}^{-1}(\Phi,\theta,\psi)V_b = T_{BE}^{-1}(\alpha_e)V_b\,, \quad (1)$$

$$\dot{V}_b = M_b^{-1}\left[F_{cg} - \omega_b \times (mV_b)\right], \quad (2)$$

$$\dot{\omega}_b = I_n^{-1}\left[M_{cg} - \omega_b \times (I_n\omega_b)\right], \quad (3)$$

$$\dot{\alpha_e} = E^{-1}(\Phi,\theta)\omega_b = E^{-1}(\alpha_e)\omega_b\,. \quad (4)$$

where:

T_{BE} – matrix of transformation from the earth frame to the body frame
P_e – position vector in regard to earth frame
V_b – vector linear velocities regarding to the body frame (u, v, w)
ω_b – vector of angular speeds regarding to the body frame (p, q, r)
α_e – vector of body orientation regarding to the earth frame, components Euler's angles (Φ, Θ, ψ)
E – transformation matrix of angular speeds
M_b – diagonal matrix of dimension 3x3, where the elements are created by total mass of the body
I_n – matrix 3x3 representing distribution of the mass along particular axes: its elements consist of moments of inertia relative to particular axes
F_{cg} – vector of resultant forces reduced to the centre of gravity of the body
M_{cg} – vector of resultant torques reduced to the centre of gravity of the body

These equations are valid for an arbitrary body with 6 degrees of freedom situated in the space. The vector α presents the body orientation in respect of our reference coordinate system that is identical with the Earth coordinate system. Resultant forces F_{cg} and resultant torques M_{cg} presents external forces/torques created by the forces and torques generated by actuator of mechatronics system. In our case it is by the main rotor and tail rotor of the helicopter. The most important vector for navigation is the vector P_e presenting the position of the body in the environment.

III. Hardware Design

The main computing unit is created by a small embedded computer based on a motherboard mini-ITX with the processor Intel i3. Advantage of the motherboard consists in integrating the supply unit with the board, so to supply the motherboard it is enough to connect a stabilized voltage 19 V with the tolerance of ±10 %. The supply source consists of five li-pol accumulators. The control computer communicates with other subsystems through USB interfaces that are converted into the RS422 bus. To eliminate vibrations during flight, the motherboard is fixed by small shock absorber blocks (Fig. 3).

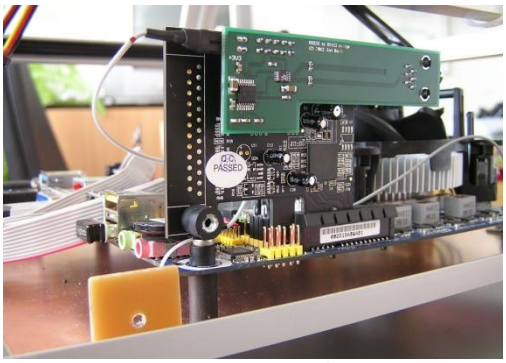


Fig. 3 Fixing the motherboard to the helicopter frame by shock absorbers

Fig. 4 shows the block interconnection of particular subsystems and control computer. Four printed boards with electronics that fulfill individual tasks are connected to the motherboard.

The main board presents a communication, supply, and control node. Embedded linear stabilizers ensure supply of servomotors for cyclic control of the helicopter; further they ensure supply of components of the board itself and they also supply other parts of the sensoric system. To the board there are connected USB busses from the control computer that are converted into the signals of the RS422 form. RS422 is then connected to the connectors of the type RJ-11.

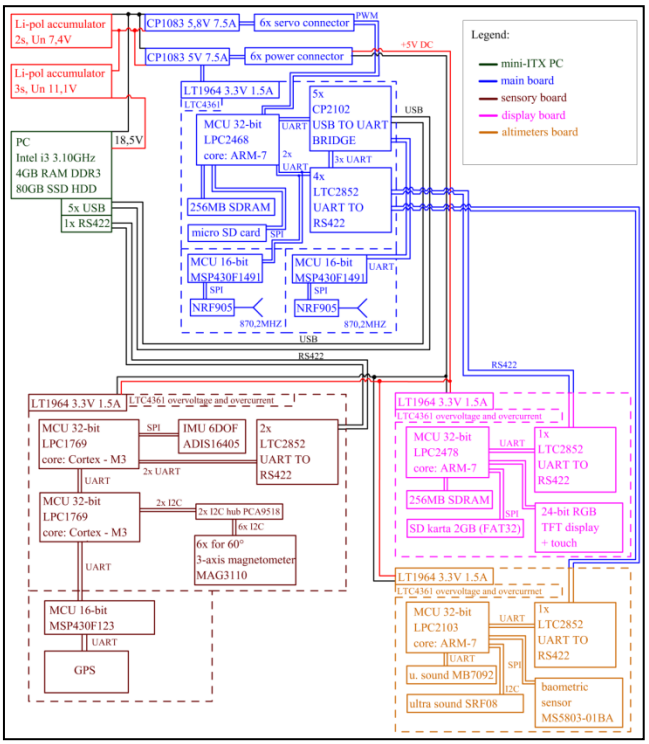


Fig. 4 Interconnection of subsystems in the control system

Communication with the operator is ensured by small high-frequency transceivers NRF905 embedded on small of printed circuit boards that are connected to the main board by connectors (Fig. 5). The processor accepts control signals from the operator and then it generates the control signals for the servomotors of cyclical control.

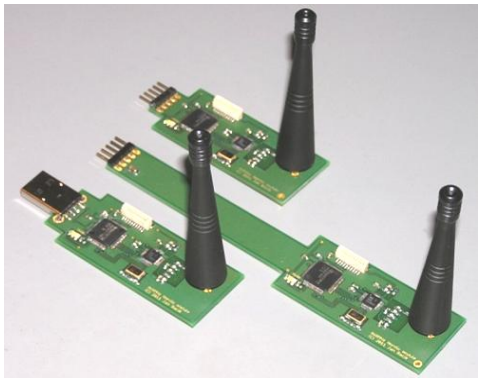


Fig. 5 Communication modules (high-frequency transceivers)

A small 24-bit TFT display embedded by a resistant touch panel creates graphical user interface (GUI). Observing it the operator sees basic data the about helicopter state and he is able to set up file name and storage site for collected data during the helicopter flight.

Among important subsystems for model identification based on experimental data there are two boards: a board of artificial horizon and board of altimeter. The main sensor of the altimeter horizon board is the sensor ADIS16405, that presents a complete inertial systems including triaxal magnetometer, triaxial accelerometer, and triaxal gyroscope with digital data output. The board is also embedded by sextuplet of triaxial magnetometers of the MAG3110 type (their arrangement is shown in Fig. 6) that serve to real-time computation of calibration data for the magnetometer installed in the sensor ADIS16405. The calibration data are computed by fitting ellipsis of the magnetic field vector trajectory. Due to presence of ferromagnetic materials in close surrounding of the sensors an ideal circular vector trajectory is deformed to an elliptical one (it is so called „hard and soft iron effect“).

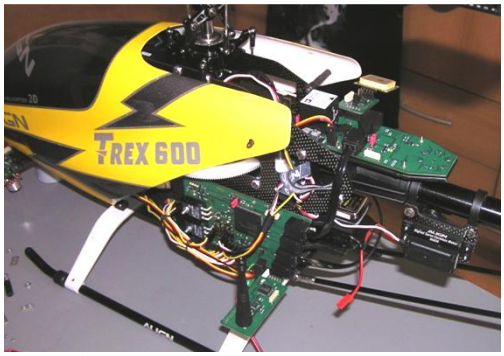


Fig. 6 Placement of sextuplet of calibration magnetometers on the board

The last sensor installed on the artificial horizon board there is a GPS receiver. The altimeter board includes ultrasound sensors by using which one can estimate helicopter model speed in the z-axis. These sensors are fixed on a two-axis gimbal ensuring a continuous direction of sound waves vertically to the earth. For flight levels determination and filtering out a terrain ruggedness there serves a small pressure sensor.

Data from gyroscope serves as information about angular rates of the helicopter and its orientation towards the earth is based on Euler angles as follows:

$$\omega_{mesured} = \omega_{real} + B_{(t)} + v \,. \tag{6}$$

The data are disturbed by a nose and time-variable bias that bring an error in correct computation of required variables. This is a reason why for the proper data processing the higher harmonics components are filtered by a low-frequency filter. Estimated Euler angles are computed as follows

$$\begin{bmatrix} \widehat{angle}_{k+1} \\ \widehat{bias}_{k+1} \end{bmatrix} = \begin{bmatrix} 1 & -dt \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \widehat{angle}_{k} \\ \widehat{bias}_{k} \end{bmatrix} + \begin{bmatrix} dt \\ 0 \end{bmatrix} \omega_E \,, \tag{7}$$

$$y = [1 \quad 0] \begin{bmatrix} \widehat{angle}_{k} \\ \widehat{bias}_{k} \end{bmatrix} + v. \tag{8}$$

Here the measured output vector y for the x and y axes presents the angles calculated from the raw accelerometer data using gravitation vector filtration:

$$\Phi = atan2(g_{Y_b}, g_{Z_b}) \,,$$
$$\theta = atan2(-g_{X_b}, g_{Z_b}) \,.$$

The measured output y for the z axis there is the angle calculated from magnetometer data:

$$\psi = atan\left(\frac{mag\_Y_e}{mag\_Z_e}\right) . \tag{9}$$

The linear speeds u, v, w are estimated by the triaxial accelerometer output signals. From the accelerometer equation

$$a_{meas.} = g + a_{real} + B + SFgcos\varphi + K(gcos\varphi)^2 + v \tag{10}$$

it follows that the accelerometer output except of the bias, scale factor, and own noise is also influenced by a gravitation force component g. Due to this reason it is necessary to filter out this component of acceleration. The filtering runs in two steps:

In the first step the data from accelerometer are re-counted from the body frame to the earth frame using the rotation matrix that contains Euler angles estimated by the discrete Kalman filter. From the data there is subsequently subtracted the gravity acceleration vector having the components [0, 0, g]

$$accl\_data_e = T_{BE}^{-1} accl\_data_b \,, \tag{11}$$

$$\begin{bmatrix} a_x \\ a_y \\ a_z \end{bmatrix}_E = \begin{bmatrix} accl\_data_x \\ accl\_data_y \\ accl\_data_z \end{bmatrix}_E - \begin{bmatrix} 0 \\ 0 \\ g \end{bmatrix}_E . \tag{12}$$

In the second step of the procedure the vector of linear acceleration vector is transformed into the body frame from reason of expressing the gravity acceleration vector in the body frame that we want to filter:

$$\begin{bmatrix} a_x \\ a_y \\ a_z \end{bmatrix}_B = T_{BE} \begin{bmatrix} a_x \\ a_y \\ a_z \end{bmatrix}_E , \begin{bmatrix} g_x \\ g_y \\ g_z \end{bmatrix}_B = \begin{bmatrix} accl\_data_x \\ accl\_data_y \\ accl\_data_z \end{bmatrix}_B - \begin{bmatrix} a_x \\ a_y \\ a_z \end{bmatrix}_B . \tag{14}$$

Acquirement of the speed vector components runs on basis of estimation of these values by a similar discrete Kalman filter like that one used for estimation of the Euler angles:

$$\begin{bmatrix} \widehat{speed}_{k+1} \\ \widehat{bias}_{k+1} \end{bmatrix} = \begin{bmatrix} 1 & -dt \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \widehat{speed}_{k} \\ \widehat{bias}_{k} \end{bmatrix} + \begin{bmatrix} dt \\ 0 \end{bmatrix} a_E \,, \tag{15}$$

$$y = [1 \quad 0] \begin{bmatrix} \widehat{speed}_{k} \\ \widehat{bias}_{k} \end{bmatrix} + v \,. \tag{16}$$

Here, as the measured output y there serves: in the x and y axis the speed derived by the GPS receiver and in the z axis the speed derived using ultrasound in combination with the pressure sensor.

IV. Experimental Results

To verify correctness of the hardware and the debugged program, the helicopter model was tested in the following maneuver: as the input signal we have chosen a ramp function of increasing the collective pitch angle of the main rotor blades θ_o while other control variables were kept on the zero level. During the test there was not blowing wind.

Fig. 7 shows time responses of the system, namely the control variables:

θ_{1s} – longitudinal cyclic pitch

θ_{1c} – lateral cyclic pitch

θ_o – main rotor collective pitch angle

θ_{oT} – tail rotor collective pitch angle

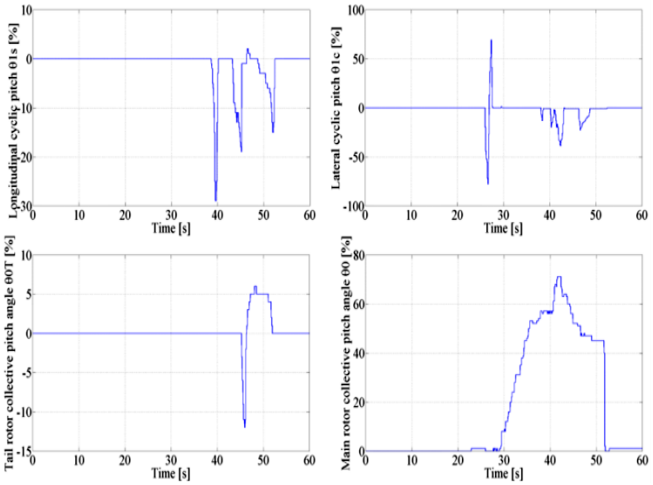


Fig. 7 Time responses of the control variables θ_{1s}, θ_{1c}, θ_o, θ_{oT}

In time t = <0 s; 30 s) the helicopter was landed on the earth and overshoots in the time courses of the lateral cyclic pitch were caused by testing establishment of connection between the helicopter and the operator. The test of response to the ramp signal θ_o was running within the time interval t = <30 s; 38 s>. During this interval the main collective pitch angle θ_o presented the only control variable. In the time t > 38 s the operator made a manual intervention there, the helicopter was stabilized, and afterwards it landed.

In Fig. 8 there are shown time responses of the helicopter state vector x (included linear speeds, angular rates and Euler angles) that corresponds to the system reaction to inputs described above. The components of angular velocities p, q, r show oscillation of the system that is substantially caused by vibration of the helicopter motor. From time responses of linear speeds u, v, w and Euler angles Φ, θ, ψ it follows that the helicopter has a tendency to lift up its nose and to move backward what is a typical phenomena of the helicopter behavior and it also try to move in direction to the right what is caused by thrust force of tail rotor.

V. Summary

The paper deals with development of hardware for a sensoric system of a small helicopter real model in order to get data from sensors of various types. After acquisition and signal data processing there follows real-time calculation of the helicopter state vector components. Designed hardware is able to process large dataflow at high sampling frequency in real time. For signal processing a discrete Kalman filter was used.

The data processing in real-time is important from view of utilization the sensoric system as a tool enabling to deliver precise data about state variables for the helicopter to the control centre. After development and verification of the sensoric system the future research will concentrate to development of a complete unmanned aerial vehicle.

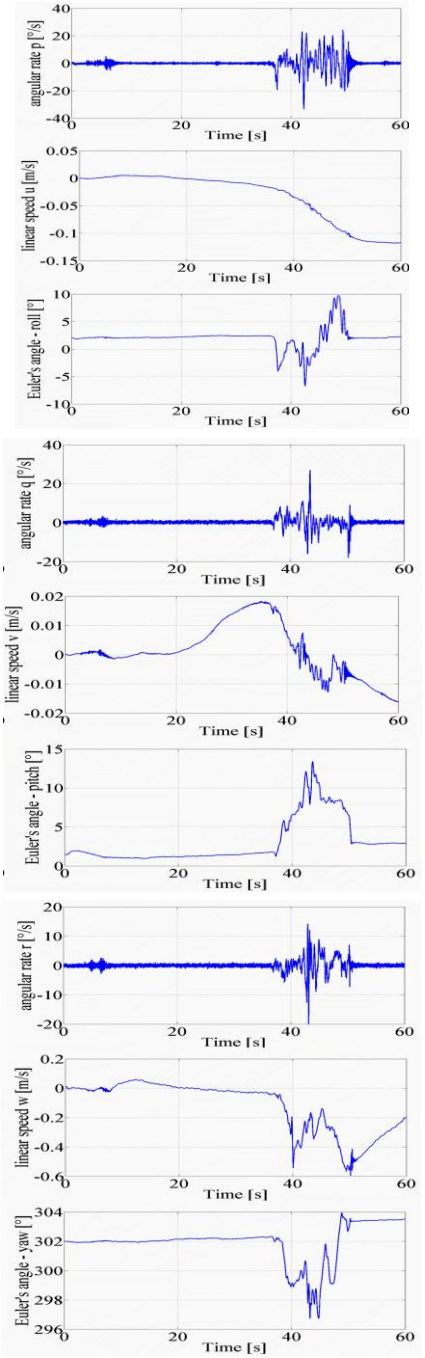


Fig. 8 Time responses of the helicopter state vector

Acknowledgment

This work was supported by the Slovak Research and Development Agency under the contract No. APVV-0185-10.

References

[1] D.H. Titterton, J.L. Weston, *Strapdown Inertial Navigation Technology*, American Institute of Aeronautics and Astronautics, Reston, 2009.

[2] T. Hsu, *MEMS & Microsystems: Design, Manufacture, and Nanoscale Engineering*, John Wiley & Sons Ltd, New Jersey, 2008.

[3] M.S. Grewal, L.R. Weill, A.P. Adrews, *Global Positioning Systems, Inertial Navigation, and Integration*, John Wiley & Sons Ltd, New Jersey, 2007.

[4] G. Cook, *Mobile Robots: Navigation, Control and Remote Sensing*, Wiley & Sons Ltd, New Jersey, 2011.

[5] M. Galassi, J. Davies, J. Theiler, B. Gough, G. Jungman, P. Alken, M. Booth, F. Rossi, *Gnu Scientific Library Reference Manual - Third Edition*, Network Theory Ltd, London, 2009.

[6] A.N. Sloss, D. Symes, Ch. Wright, *ARM System Developer's Guide. Designing and Optimizing System Software: Designing and Optimizing System Software*, Morgan Kaufmann Publisher, New York, 2007.

Device for pulse annealing of amorphous magnetic ribbon

[1]*Lukáš HUBAČ (1st year)*
Supervisor: [2]Ladislav NOVÁK

[1]Dept. of Physics, FEI TU of Košice, Slovak Republic
[2]Dept. of Physics, FEI TU of Košice, Slovak Republic

[1]lukas.hubac@tuke.sk, [2]ladislav.novak@tuke.sk

***Abstract*— Present article describes extension of the existing apparatus for measurements of magnetic properties of amorphous ribbons. This upgrade apparatus allows the pulse annealing of ribbons. Realization of this device consists of the design and construction of special non-magnetic holder for the sample attachment. It also includes design and construction of pulse source. In addition, also the control program was extended by the part for controlling of the pulse heating device. . Experiment was realized on a sample called FINEMET. Obtained experimental data were processed.**

***Keywords* — amorphous ribbon - FINEMET, pulse heating, measurement of magnetic induction**

I. INTRODUCTION

Soft magnetic materials used as cores of transformers and parts of electric motors are characterized by low magnetic reversal losses. So, they have a narrow hysteresis loop resulting in low coercivity (less than 100 A/m). Ferromagnetic amorphous alloys in the form of ribbons belong to these materials. They are prepared by rapid cooling of the melt. They are sensitive to the magnetic field and they have a large tensile strength. Therefore, they provide a variety of applications. They are ideal as magnetic field sensors, tensile force sensors [1]. They are used as kernels of small transformers due to their good magnetic properties and small dimensions. A conventional transformer metal sheet can be replaced by them.

The basic principle of production of the amorphous ribbons from the melt is very rapid transformation melt to thin layer and its subsequent rapid cooling (10^6 K/s) [1]. Finally a very thin film or ribbon is produced. This process can be realized by various techniques: hitting droplets of the melt on the cold metal surface, compression of droplets between two plane surfaces or slinging droplets of the melt between two rotating cylinders. The most sophisticated method is casting a thin stream of melt on the circumference rotating copper wheel. In this way we can prepare ribbons with thickness of a few μm and with width of a few cm [2].

Magnetic properties of ferromagnetic amorphous alloys can be improved by heat treatment. Alternatively, the properties can be also improved by heat treatment with simultaneous application of magnetic field and/or of mechanical stress. [3]. Application of mechanical stress during process of annealing leads to the increase of induced anisotropy of amorphous ribbon. Annealing reduces internal tension and so improves magnetic properties. Processing at higher temperatures may also induce the internal stress as a result of rapid heating and cooling or the onset of crystallization.

In presented work, an assumption that the magnetic soft material can be improved by application of combined impulse - annealing with application of mechanical stress - is studied. The reason is that pulse annealing can prevent crystallization at very high temperatures exceeding the Curie temperature (T_c). The crystalline material is very fragile and breakable. This is a disadvantage of nanocrystalline materials. Classic conventional heating, by at annealing at T_c would also lead to nanocrystallisation (diffusion over a long distance). The role of the pulse heating is to achieve diffusion only on short distance, so there is only a rearrangement of atoms, no crystallization.

II. EXPERIMENT

For the study of pulse heating material known as FINEMET ($Fe_{73,5}Si_{13,5}B_9Nb_3Cu_1$) prepared in MTA KFKI in Budapest has been selected. Amorphous samples and nanocrystalline samples after heat treatment at 800 K for 30 minutes were used. The width of the samples was about 10 mm, thickness 30 μm and length 10 cm. Hysteresis loops of both samples were measured in detail and then some magnetic parameters have been specified.

Table 1 shows the selected magnetic parameters of samples: coercivity (H_C), magnetic induction saturation measured in the field 15 kA/m (B_s) and Curie temperature (T_C).

TABLE 1
SELECTED MAGNETIC PARAMETERS OF THE SAMPLES

Sample	H_C (A/m)	B_S (T)	T_C (K)
FINEMET amorphous	17,1	1,180	600
FINEMET nanocrystalline	2,04	1,185	800

Hysteresis loops of both samples are shown in Figure 1. Details of measured loops in low field region are shown in Figure 2.

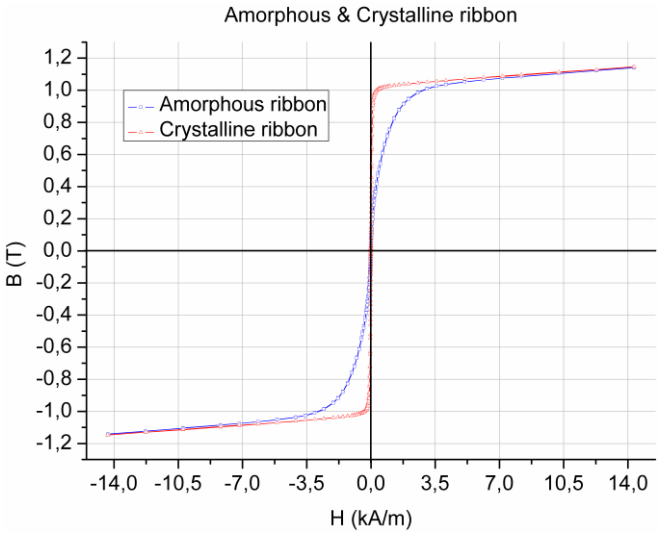


Fig. 1. Hysteresis loop of amorphous and nanocrystalline samples FINEMET.

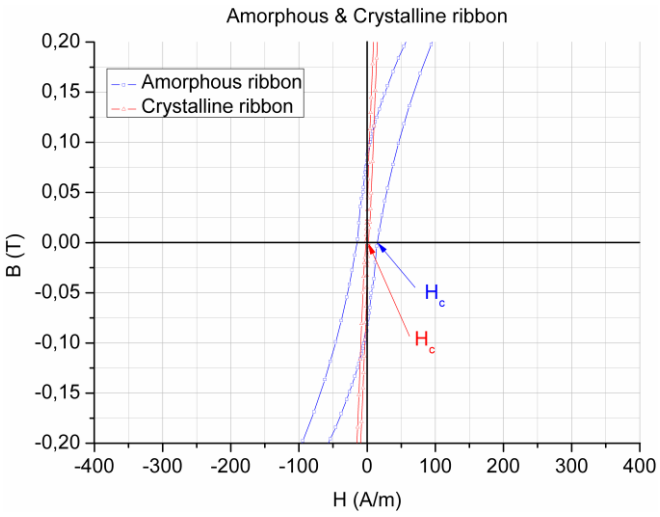


Fig. 2. Detail of the hysteresis loop of amorphous and nanocrystalline samples FINEMET.

From these results follow that nanocrystalline transformation improves quality of magnetic soft materials. The undesirable property of these materials is that they are very fragile.

In this article, the problem of heat treatment of amorphous ribbons is solved by applying a pulse heating. . The major priority is to achieve a significant improvement in the magnetic properties while maintaining the amorphous state.

III. EXPERIMENTAL DEVICE

The task of this item is to describe the existing part of apparatus for measuring the magnetic properties of soft magnetic material which was constructed by me. It is a device for the implementation of pulse heating of a sample directly in the measuring device. This method has some advantages for the investigations of results of the pulse annealing.

For example:

- The duration of the pulse can be easily controlled in a wide time range from 0,02s to whole seconds, so the temperature is very precisely controlled.
- Measurement of magnetization can be realized immediately after annealing in the same apparatus.

The sample is placed vertically in vertically oriented solenoid with a diameter of 40 mm and a height of 500 mm in special holder from top part. This mount is made of non-magnetic material and is arranged so that the sample could be annealed and also burdened with a variety of vertical forces in tension. The sample can be annealed in the magnetic field of solenoid. The detail of the holder is shown in the figure 3.

Fig. 3. View on holder and stainless steel cover.

Apparatus for pulse annealing is composed of a special holder and control part. The control part consists of:

- computer-controlled pulse source
- part for scanning of pulses
- control program

The source consists of control part and power part, which are electrically isolated from each other by an optocoupler (photo-diac). Power section consists of a power triac and transformer, whose transformation ratio is approximately 0,1 and power 250W. The current in the primary winding of the transformer is controlled by triac that is switching photo-diac (MOC 3041) in the zero. Optical element is controlled by computer with installation laboratory card ADLINK ACL-8316 by TTL logic. The secondary winding of the transformer

is connected directly to the sample. The current (about 10A) passes through the sample and the Joule heating occurs.

In order to accurately control the number of pulses, it was necessary to do further extension of apparatus for feedback. The output of this feedback in the form of TTL logic is sensing transition AC zero. This signal is fed to the control computer with program. This program manages the entire Joule heating and the subsequent measurement of the magnetic properties of the sample. Wiring diagram of pulse source and the sensor transition AC zero, is shown in Figure 4.

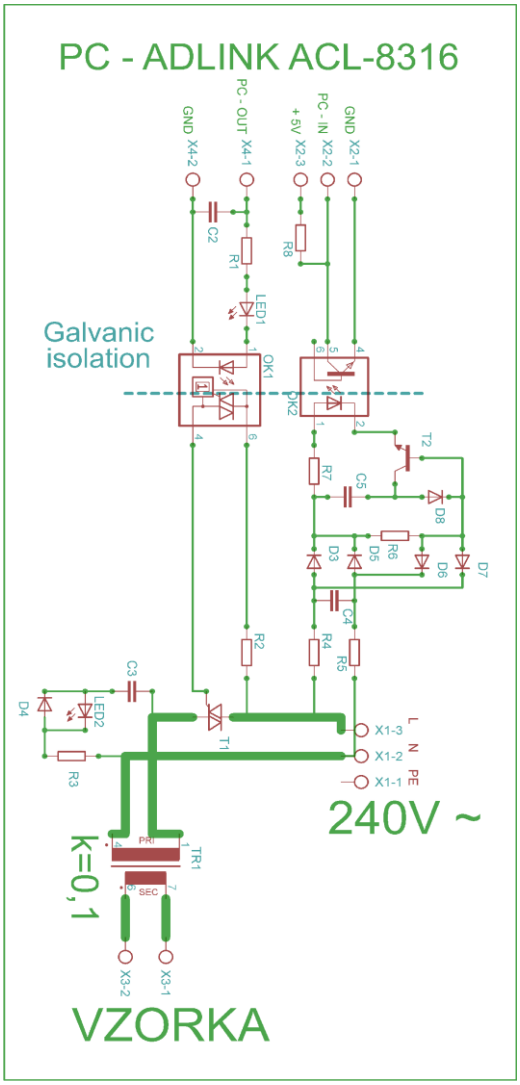


Fig. 4. Wiring diagram of pulse source and the sensor transition AC zero.

Impulse heating and magnetization measurements were realized in the same sample holder. Impulse heating was realized by alternating the current pulses of approx. 10A, 50Hz. Pulse duration was 0,3 - 0,5 s. Samples were processed in this heat without forced air cooling. So, the estimated cooling rate is about 100 K/s. Thermal heating rate was about 1000 K/s.

IV. Conclusion

There was designed and constructed special non-magnetic holder for sample attachment. Moreover, the pulse source for sample annealing was designed, and subsequently constructed. Furthermore, the control program was programmed. Finally, there were processed test measurements and analysis of obtained experimental data.

Acknowledgment

The author would like to acknowledge his supervisor doc. RNDr. Ladislav Novák CSc. for proposing the topic of this article and his continuous help.

References

[1] P. Ripka, *Magnetic Sensors and Magnetometers*, Artech House, Boston, London, 2001.
[2] H. H. Liebermann, L .D. Graham, IEEE Trans. Magnet., MAG-12, 6(1976), p.921.
[3] L. A. Davis, N. DeCristofaro, C. H. Smith, Proc. Conf. Metallic Glasses: Science and Technology, Budapest, vol. 1, 1980, p.1.

Modeling the distribution of electromagnetic field and influence of shielding material

[1]*Ján Zbojovský (2st year),* [2]*Marek Pavlík (2st year)*
Supervisor: [3]*Alexander Mészáros*

[1,2,3]Dept. of Electric Power Engineering, FEI TU of Košice, Slovak Republic

[1]jan.zbojovsky@tuke.sk, [2]marek.pavlik@tuke.sk, [3]alexander.meszaros@tuke.sk

***Abstract*— In this paper is described the simulation of electromagnetic field and influence of some materials to the its distribution. For simulation copper, iron and aluminium material is chosen, which is used as a shielding material.**

***Keywords*—electromagnetic field, simulation, shielding**

I. Introduction

The issue of the impact of electromagnetic fields on the population began to be actual due to a significant increase the resources of the electromagnetic field caused by the increasing electricity consumption as well as the rapid development of telecommunication technology. At present it is not possible to confirm or deny possible harmful effects of electromagnetic fields on human health. The significant number of completed and published research says, that this issue is actual. To protect themselves and the public is an important EMC. It is the ability to present proper function (or coexistence) equipment or systems - electrical and biological - located in a common electromagnetic environment without significant adverse influence on their normal functions. Electrical systems must be resistant to other systems and their action may adversely affect the normal operation of other systems and devices. Electromagnetic interference can cause severe problems, and it should be taken into account in the design of new electrical devices [1], [2].

II. Effect of the Electromagnetic field and shielding coefficient

The electromagnetic field is a physical field in which electric and magnetic forces operate in space. Nowadays, the effects of electromagnetic fields on living organisms are discussed by the International Organization for Non-Ionizing Radiation Protection (INCIRP) and the World Health Organization (WHO). Biological effects are measurable changes induced by some stimulus or change in the environment. These changes do not necessarily lead to damaging health. The human body has many compensatory mechanisms and the ability to adapt to various changes. These reactions are normal signs of life. Biological effect, however, could have an adverse effect in the case where the body is exposed to long periods of harmful factors and compensatory mechanisms are insufficient [3].

One of the basic ways to protect against the effects of electromagnetic fields is shielding of equipments.

Shielding is an important part of the equipment both in terms of mutual interference, and to protect the public against electromagnetic radiation. Properties of shielding can be presented by the coefficients of shielding, absorption and reflection.

Shielding coefficient is determined by this relation:

$$K_s = \frac{H_2}{H_1} \tag{1}$$

where H_1 is the magnetic field intensity impinging on the shield barrier and H_2 represents the intensity of the magnetic field at some point of shielded area.

Shielding effectiveness is expressed by:

$$SE = 20 \log \frac{1}{K_S} \tag{2}$$

Based on the IEEE standard Shielding effectiveness is determined for the different frequency bands as follows [4], [5]

The frequency range from 50Hz to 20MHz:

$$SE = 20 \log \frac{H_1}{H_2} (\mathrm{dB}) \tag{3}$$

or

$$SE = 20 \log \frac{V_1}{V_2} (\mathrm{dB}) \tag{4}$$

where V_1 is the voltage the shield barrier and V_2 represents the voltage at some point of shielded area.

For the frequency range from 20MHz to 300MHz-100GHz the relations are following:

$$SE = 20 \log \frac{E_1}{E_2} (\mathrm{dB}) \quad (5)$$

or

$$SE = 20 \log \frac{P_1}{P_2} (\mathrm{dB}) \quad (6)$$

where P_1 is the power impinging on the shield barrier and P_2 represents the power at some point of shielded area.

The shielding effectiveness is the sum of absorption and reflection:

$$SE = A + R (\mathrm{dB}) \quad (7)$$

where A is the absorption coefficient, B is the reflection coefficient. Absorption coefficient is determined by:

$$A = 8{,}69 \cdot \frac{t}{\delta} (\mathrm{dB}) \quad (8)$$

where t is the thickness of shielding barrier, δ is the depth of penetration of the electromagnetic field in the material, which is determined by:

$$\delta = \sqrt{\frac{2}{\omega \mu \sigma}} (\mathrm{dB}) \quad (9)$$

Where μ is the permeability and σ conductivity of shielding material.

Reflection coefficient is determined by:

$$R = 20 \log \left(\frac{1}{4} \cdot \sqrt{\frac{\sigma}{\omega \mu_r \varepsilon_0}} \right) (\mathrm{dB}) \quad (10)$$

where ε_0 is the vacuum permittivity and μ_r permeability of shielding material.

III. Modeling the Electromagnetic Field

For the numerical calculation of the electromagnetic field distribution there are used Maxwell's equations that are supplemented by the material equations. For their solution it is appropriate to use some numerical methods (finite element method (FEM), finite-difference method (FDM) and boundary element method (BEM)).

Professional computer programs that use the FEM for solving electromagnetic fields are suitable for their versatility and also for this type of task. The presented model of poles was simplified into 2D space, so the calculation time was shorten compared to the 3D model [8].

Numerical Methods

The essence of numerical methods is the discretization of continuous variables, which means that the differential equations are expressed as a set of algebraic equations. Their target is to find a solution in the final examination process, the finite number of discrete points.

A common feature of numerical methods is the replacement the exact solution differential or integral equations by their approximate solution.

Each task is specified by the following parameters:

• describing the geometry of the shape and size of the object of interest. The area in which is searching the solution, is divided to sub-areas and those are divided into elements, nodes and their coordinates.

• a description of the material properties of sub-environments.

• description the distribution of the field sources and their properties in sub-areas and on surfaces,

• differential or integral equation of the field,

• a description of the boundary conditions at the border sub-regions.

• other relationships derived for the calculation of secondary parameters.

This problematic is discussed in the literature [6], [7].

IV. Simulation Program and Modelling

For the simulation was chosen ANSYS. This program is a software package based on the finite element method. It is intended for solving large-scale linear and non-linear tasks many categories: structural, thermal, thermal - mechanical, electromagnetic, acoustic, etc.

It will simulate the distribution of the electromagnetic field and the impact of various barriers to their distribution. The basic procedure for the simulation can be summarized in the following four points.

• Initial decision:
  First, it is necessary to determine what type of analysis is (static, dynamic, linear, etc.). Then what will be modeled, and finally how it is modeled.

• Preparation of the model (Preprocessing)
  This includes the type of elements, assign material properties, import or creation of the geometric model, creating a finite element network and determination of boundary conditions.

• Solution

• Evaluation and check of results

Description of the simulation

Source of the electromagnetic field and the environment that surrounds the source will be simulated. In this environment (e.g. air) will be placed shielding material (iron, aluminium and copper alloy) and we will be able to see how these materials affects the distribution of the electromagnetic field. Based on the obtained results are determined the shielding coefficients.

Parameters of simulation

The diameter of a source is 2.24 mm, distance the barrier from the source is 1 mm and the dimensions of the barrier are 1x20 mm. Environs are represented by air.

TABLE I
Material properties

Material	Isotropic Relative permeability	Isotropic resistivity (Ωm)	Temperature (°C)
Iron	10000	$1.7*10^7$	20
Copper alloy	1	$1.694*10^8$	20
Aluminum alloy	1	$2.67*10^8$	20

In the first case copper alloy was chosen.

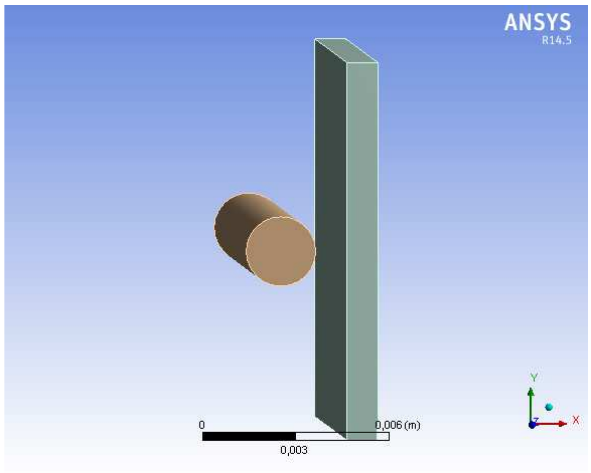


Fig. 1. Model the source and barrier

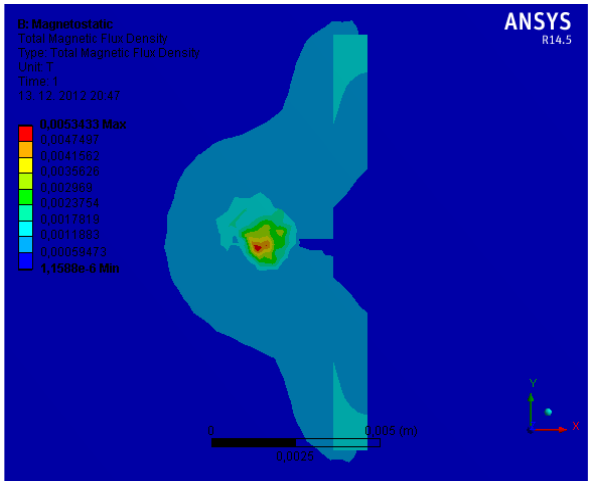


Fig. 2. The distribution of the magnetic field in space

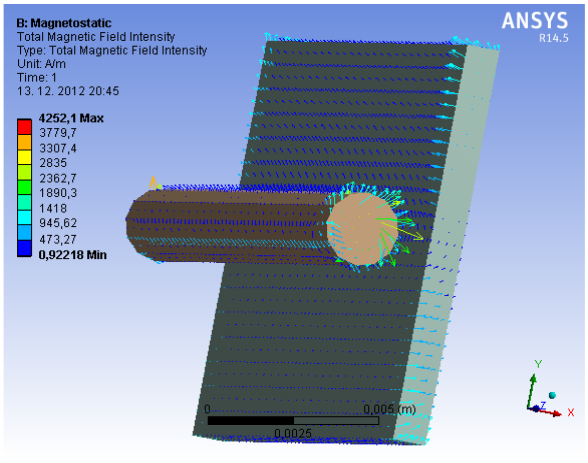


Fig. 3. The vector distribution of the magnetic field in space

For the next case of simulation iron was chosen.

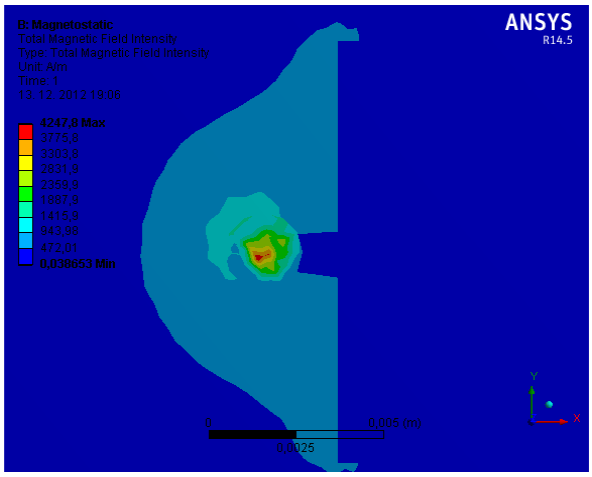


Fig. 4. The distribution of the magnetic field in space

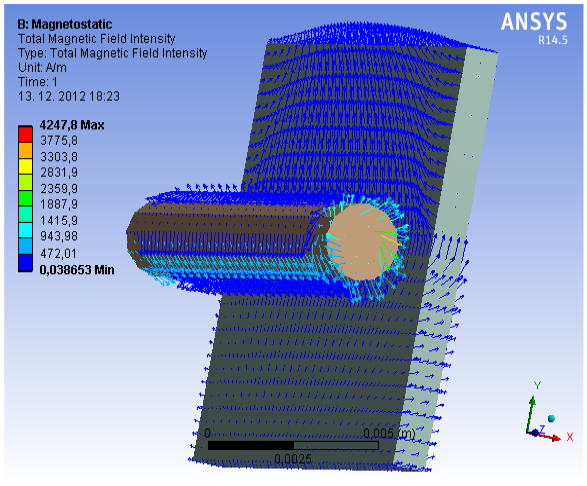


Fig. 5. The vector distribution of the magnetic field in space

And the last case was represented by aluminium alloy.

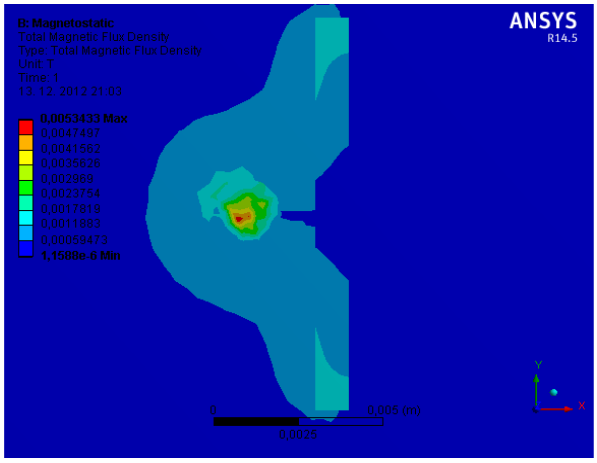


Fig. 6. The distribution of the magnetic field in space

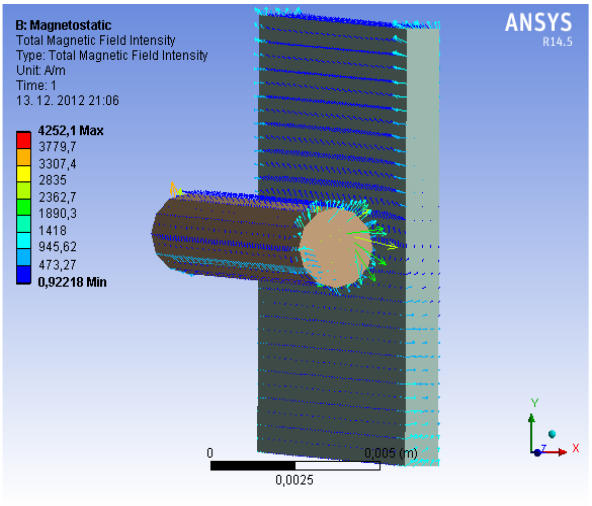


Fig. 7. The vector distribution of the magnetic field in space

TABLE 2
DATA FOR CALCULATING THE COEFFICIENTS

Material	Intensity of the field on the barrier [A/m]	Intensity of the field behind the barrier [A/m]	Shielding effectiveness [%]
Iron	600,69	20,14	29,49
Copper alloy	331,6	145,41	7,160
Aluminum alloy	557,23	139,81	12,01

V. CONCLUSION

Nowadays the problematic of the electromagnetic field began to be very actually, and it is important to be engaged in this topic. One of the ways to solve the distribution of the electromagnetic field is measurement or simulation and in this paper simulation is presented. The influence of some materials to the distribution of the magnetic field was simulated. All the results are summarized in table 1 and on fig.1-7 is displayed the distribution of magnetic field. According to the relations for shielding effectiveness were calculated that the best shielding effectiveness have the iron.

ACKNOWLEDGMENT

This publication is the result of the Project implementation: Protection of the population of the Slovak Republic from the effects of electromagnetic fields, ITMS: 26220220145, Activity 2.1 supported by the Research & Development Operational Programme funded by the ERDF.

Európska únia
Európsky fond regionálneho rozvoja

Operačný program VÝSKUM a VÝVOJ

Agentúra Ministerstva školstva, vedy, výskumu a športu SR pre štrukturálne fondy EÚ

We support research activities in Slovakia / Project is cofinanced from EU funds.

REFERENCES

[1] Mészáros, A.: Elektrotechnika a problémy životného prostredia. Košice, KEE FEI TU, 2009. 304 p. ISBN 978-80-553-0175-4

[2] Kmec M., Hvizdoš M., Skúšky digitálnych ochrán prístrojom Omicron CMC, Electrical Engineering and Informatics 3: proceeding of the Faculty of Electrical Engineering and Informatics of the Technical University of Košice. - Košice: FEI TU, 2012. p. 705-710. ISBN 978-80-553-0890-6

[3] Vecchia P, et al: Exposure to high frequency electromagnetic fields, biological effects and health consequences (100kHz to 300GHz), INCIRP 16/2009

[4] Űnal E., Gőkcen A., Kutlu Y.: Effective Electromagnetic Shielding, IEEE Microwave magazine, 1527-3342, 2006, p. 48 – 54

[5] Rusiecki Andrzej: Calculations and measurements of shielding effectiveness of slotted enclosure with built-in conductive stirrer In: Przegland Elektrotechniczny, ISSN 0033-2097, R. 88 NR 10b/2012, p. 328-329

[6] Dědková Jarmila: Modelování elektromagnetických polí, Vysoké učení technického v Brně, 2006, 82 p.

[7] Mayer, D., Ulrych B.: Základy numerického řešení elektrických a magnetických polí, SNTL/ALFA Praha 1988, 208 p., 04-528-88

[8] Medveď, D., Modeling of Electromagnetic Fields Close to the Very High Voltage and Extra High Voltage Poles In Elektroenergetika, Vol.5, No. 2, 2012. ISSN 1337-6756, p. 17-19

Predictive Power Control of Single-phase Grid-tied Cascade Inverter

[1]*Marek PÁSTOR (3rd year),* [2]*Mišél BATMEND (3rd year)*
Supervisor: [3]*Jaroslav DUDRIK*

[1,2] Dept. of Electrical Engineering and Mechatronics, FEI TU of Košice, Slovak Republic

[1]marek.pastor@tuke.sk, [2]misel.batmend@tuke.sk. [3]jaroslav.dudrik@tuke.sk

***Abstract*—The paper presents a power control scheme for grid-tied multilevel cascade inverter with modern predictive current control technique. The reference current for current predictive controller is generated using instantaneous power theory (p-q theory). The behavior of the system is simulated for different grid current distortions. The quality of proposed control scheme is evaluated as p-q theory was originally designed for three-phase active power compensators and can not be considered as general power theory. The main advantage of proposed control scheme is its simplicity. Due to the use of predictive controller the control structure can be designed in stationary reference frame without the need of resonant current controller and PWM modulator.**

***Keywords*—cascade inverter, p-q theory, predictive control.**

I. Introduction

The power electronics plays an important role in today's renewable energy systems. The electrical energy harvested from a photovoltaic (PV) generator needs to be conditioned in order to by supplied to the electrical grid by a mean of PV inverter.

The modern PV inverter can be considered to by a mature technology. However there are still several problems which need to be addressed. Some of them are higher order harmonics in supplied ac current, common voltage reduction, EMC, output filter size, reliability and lifetime, etc. In order to solve these problem there are several possibilities. One can use modern semiconductor components such as SiC, higher switching frequency, soft-switching technique, new topology of power converter, etc. One of the possibilities is to use multi-level converters. The main advantages of multilevel converters when compared to dual-level converters are staircase output voltage with lower du/dt, lower THD, smaller power transferred by one semiconductor switch, lower switching losses. Some of the abovementioned problems can be solved by incorporating modern control techniques such as fuzzy logic control, neural networks or predictive control. Traditionally, control technique had to ensure fast dynamics and stability of the systems. Nowadays requirements such as switching losses minimization, lower THD good performance in wide range of operating conditions, considering prohibited switching states, active dumping of oscillations, etc. are taking into account as well. [1][2]

When considering modern power converter with finite number of switching states, nonlinearities and constrains and modern control implementation in discrete-time, known model of power converter and availability of fast control platforms the predictive control technique comes as natural solution.

Most advantages can be gained by combining modern multilevel converter with modern predictive control strategy. The scope of this paper is to describe predictive control of cascade multilevel converter used for PV inverter with emphasis on instantaneous grip-supplied power control.

II. Grid Connected Cascade Inverter

A. System Description

The PV generator is connected to the grid by a mean of single-phase cascade H-bridge inverter with output grid filter as shown in Fig. 1. The current control of the inverter is done by predictive control technique which ensures high dynamics. The reference current for predictive controller is generated by power control block which accepts required active and reactive power supplied to the grid as inputs. The control is realized in stationary reference frame.

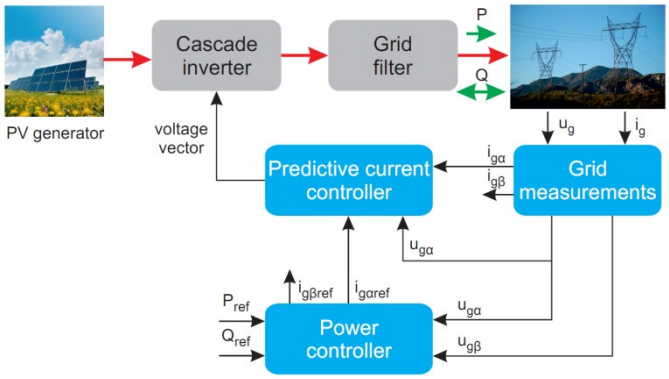


Fig. 1. Proposed control scheme for power control flow of grid connected cascade inverter

B. System Model

The main feature of predictive controller is the use of system model to predict the future behavior of controlled variables. The controlled variable in grid connected inverter with current control is the current supplied to the grid. The amplitude and phase shift of grid current with respect to the

grid voltage are defined by required active and reactive power supplied to the grid.

The voltage source cascade inverter is connected to the grid through L filter. The system is described in discrete time by (1):

$$i_g(k+1) = \frac{T_S}{L}\left(v(k) - u_g(k)\right) + i_g(k)\left(1 - \frac{RT_S}{L}\right) \quad (1)$$

where:

i_g – is current supplied to the grid,
v – is the output voltage of the inverter,
u_g – is grid voltage,
L – inductance of L filter,
R – resistance of L filter,
T_S – sampling time.

The equation (1) is used in predictive controller to predict the future value of the grid current based on actual grid current and voltage as well as the inverter output voltage (Fig. 2). The proposed asymmetrical H-bridge cascade inverter is capable of creating 15 discrete levels of output voltage.

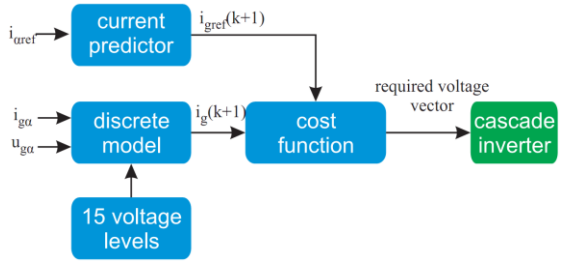


Fig. 2. Predictive current controller of cascade inverter

For predictive control, there is a need to create the cost function which will be evaluated in each sampling time and will define the behavior of the system. The cost function was chosen as difference between desired current i_{gref} and actual current i_g in the next sampling period [8]:

$$z(k) = \left|i_{gref}(k+1) - i_g(k+1)\right| \quad (2)$$

C. Grid Measurements

In order to realize a control system of the one-phase cascade inverter in stationary reference frame there is a need to create grid measurements in stationary reference frame αβ. This is normally done by Clarke transformation in three-phase systems. However, Clarke transformation can not be utilized in one-phase system. Thus there is a need to create virtual two-phase generator in order to create stationary reference frame αβ. This virtual two-phase generator needs to be used in grid voltage as well as grid current measurements. In the following, the virtual two-phase generator for voltage measurements in described. The current measurement principle in the stationary reference frame is the same and thus is not mentioned.

The property of the stationary reference frame is that two voltages v_α and v_β are orthogonal. If the grid voltage corresponds to the voltage $u_{g\alpha}$, than the voltage $u_{g\beta}$ can be created as follows:

$$\begin{bmatrix} u_{g\alpha} \\ u_{g\beta} \end{bmatrix} = \begin{bmatrix} u_g(\omega t - \pi/2) \\ u_g(\omega t) \end{bmatrix} = \begin{bmatrix} U_{gm}\sin(\omega t - \pi/2) \\ U_{gm}\sin(\omega t) \end{bmatrix} \cong \begin{bmatrix} -V_m\cos(\omega t) \\ V_m\sin(\omega t) \end{bmatrix} \quad (3)$$

There are several possibilities how to create the 90 degrees phase shift of the grid voltage to produce the $u_{g\beta}$ voltage (e.g. storage elements, filters). One of them is to use second-order low-pass filter [5][6].

When the input voltage u_g passes through the second-order low-pass filter, which damping ratio $\zeta = 1/\sqrt{2}$, the undamped natural frequency ω_n has the same value as the grid frequency, a signal with a phase-angle difference of π/2 and amplitude of $U_{gm}/\sqrt{2}$ is obtained [5]:

$$u_{g\alpha} = -\sqrt{2}\frac{U_{gm}}{\sqrt{2}}\sin\left(\omega t - \frac{\pi}{2}\right) = U_{gm}\cos(\omega t) \quad (4)$$

The second order low-pass filter is realized in discrete form with help of Tustin transformation. The resulting grid measurement block generating voltages and currents in the stationary reference frame is shown in Fig. 3.

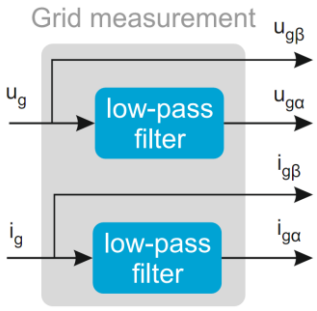


Fig. 3. Inside of grid measurement block

III. Power Control

The power supplied to the grid by the cascade inverter is defined by grid voltage and grid current. The grid voltage is considered as external variable and can not be changed. Thus the power control is realized solely by current control. By generating the required grid current based on grid voltage and required supplied active and reactive power and ensuring current control technique, the power control is realized as well. One way how to generate the required grid current to met power requirements is to use instantaneous power theory also known as p-q theory presented by Akagi in [4].

The p-q theory was originally developed for active power factor compensators. It is based on the Clarke transformation of grid voltage and current in three-phase system into αβ stationary reference frame. It defines two instantaneous powers named as instantaneous active p and instantaneous imaginary q power (with no physical meaning):

$$\begin{bmatrix} p \\ q \end{bmatrix} = \begin{bmatrix} u_{g\alpha} & u_{g\beta} \\ -u_{g\beta} & u_{g\alpha} \end{bmatrix}\begin{bmatrix} i_{g\alpha} \\ i_{g\beta} \end{bmatrix} \quad (5)$$

The p-q theory can be used for one-phase systems as well. However, the resulting instantaneous active and imaginary powers are doubled with respect to real one-phase powers:

$$\begin{bmatrix} p \\ q \end{bmatrix} = \frac{1}{2}\begin{bmatrix} u_{g\alpha} & u_{g\beta} \\ -u_{g\beta} & u_{g\alpha} \end{bmatrix}\begin{bmatrix} i_{g\alpha} \\ i_{g\beta} \end{bmatrix} \quad (6)$$

In [7] is shown that for sinusoidal balanced three-phase system is the instantaneous reactive power equal to the active power and instantaneous imaginary power is equal to the reactive power:

$$p = P$$
$$q = Q \qquad (7)$$

However for either non-sinusoidal or unbalanced condition the (6) is not valid any more. In general, the p-q theory is not a general power theory [7].

The reference current for predictive current controller can be obtained by combining (6) and (7):

$$\begin{bmatrix} i_{garef} \\ i_{g\beta ref} \end{bmatrix} = \frac{1}{u_{g\alpha}^2 + u_{g\beta}^2} \begin{bmatrix} u_{g\alpha} & -u_{g\beta} \\ u_{g\beta} & u_{g\alpha} \end{bmatrix} \begin{bmatrix} P_{ref} \\ Q_{ref} \end{bmatrix} \qquad (8)$$

Only the current i_{garef} is used as reference current for predictive current controller because the grid voltage u_g corresponds to the voltage $u_{g\alpha}$.

IV. Simulation

A. Simulation Setup

The proposed power control technique of single-phase grid-connected cascade inverter with predictive current control was verified by simulation in MATLAB/Simulink. The program realization of control scheme from Fig. 1 is shown in Fig. 4. The asymmetrical 15-level cascade inverter has DC voltage levels of 60, 120 and 240 V. The grid filter has inductance of 10 mH and resistance of 10 mΩ. The grid voltage is ~ 230V/50Hz and the sampling time T_S was set to 100 µs.

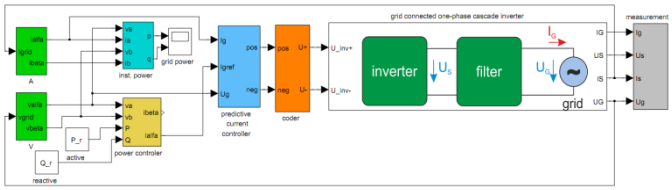


Fig. 4. Simulation scheme in Simulink

B. Instantaneous Power

At first, the accuracy of p-q theory when measuring instantaneous power in one-phase system was verified. The reason is that p-q theory is not general power theory [7] and the current supplied to the grid by the cascade inverter has some distortion. The accuracy of the p-q theory under these circumstances could be solved analytically. However, for particular inverter topology and control technique the easiest way is to use simulation.

The simulation results are shown in Fig. 5 and 6. There was a step change in active power on 0.05 and 0.2 s and in reactive power on 0.125 and 0.25 s. The active and reactive power was measured with p-q theory and Active & Reactive Power meter from SimPowerSystems toolbox which measures the power of the first harmonic. The grid filter inductance L was set to 10 and 2 mH to simulate different distortion in grid current. As can be seen from Fig. 5 and 6 the power measured by p-q theory has high-frequency oscillations but the average value is close to one measured by Active & Reactive Power meter. The advantage of p-q theory is faster response time as there is no need to calculate the rms values of current and voltage. The fast response time is desired for high-dynamic predictive control.

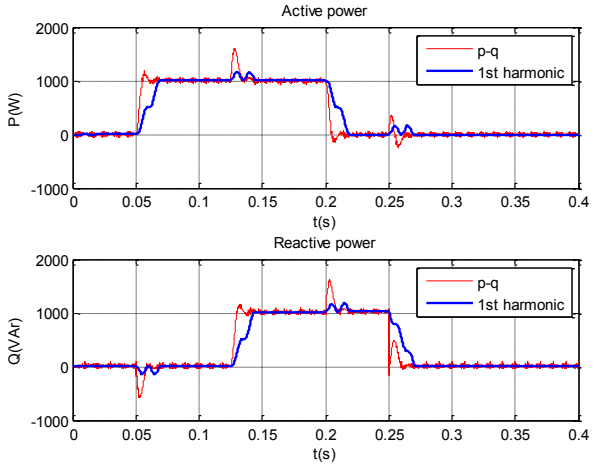


Fig. 5. Comparison of power measurement with p-q theory and first-order harmonic power meter Active & Reactive Power from SimPowerSystems toolbox for L = 10 mH

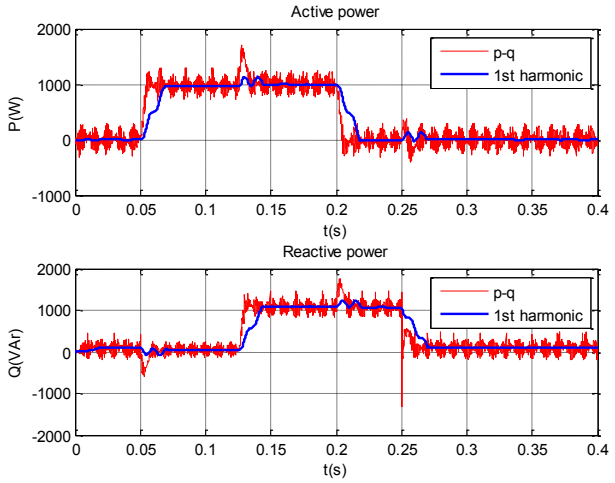


Fig. 6. Comparison of power measurement with p-q theory and first-order harmonic power meter Active & Reactive Power from SimPowerSystems toolbox for L = 2 mH

C. Power Control

The power control was realized using (8). The inputs were required active and reactive power and the output was required grid current. The quality of power control was verified for two values of grid filter inductance: 10 and 2 mH. Simulation results are shown in Fig. 7 and 8.

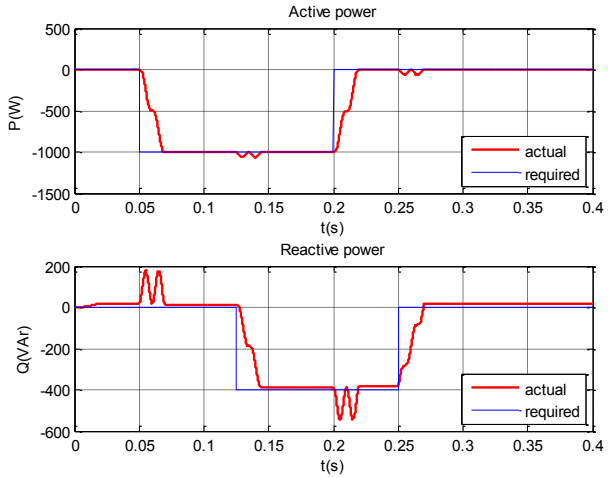


Fig. 7. Comparison of required and actual power measured with Active & Reactive Power from SimPowerSystems toolbox for L = 10 mH

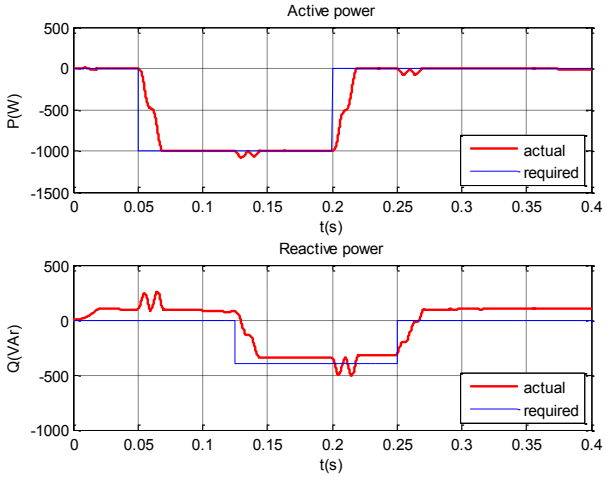


Fig. 8. Comparison of required and actual power measured with Active & Reactive Power from SimPowerSystems toolbox for L = 2 mH

Lower value of grid filter inductance causes higher grid current ripple, as indicates Fig. 10. This higher current ripple generates higher distortion power which results in imbalance between required and desired reactive power. As the value of the inductance is decreased the imbalance between required and actual reactive power is more significant (Fig. 8).

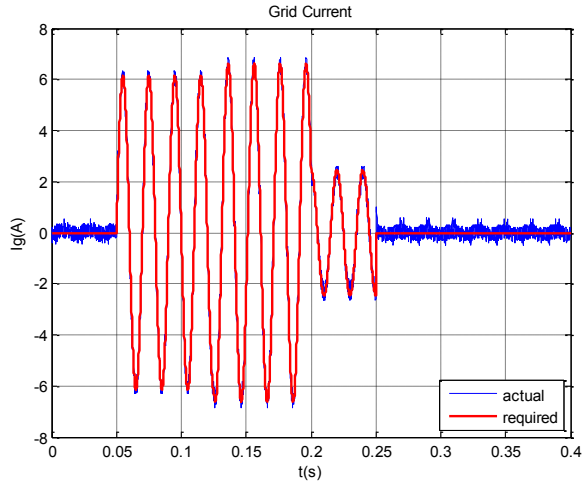


Fig. 9. Comparison of required and actual grid current for L = 10 mH.

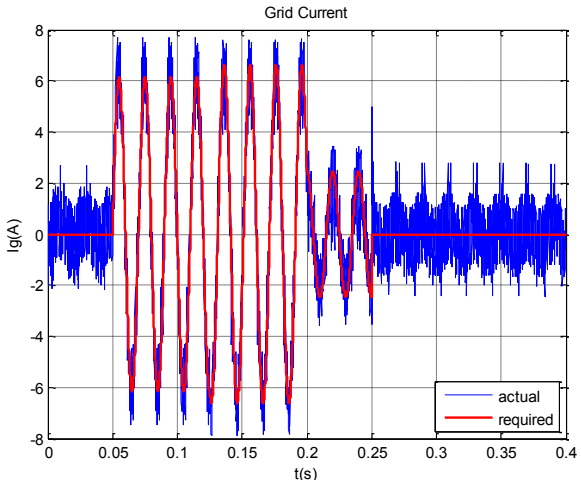


Fig. 10. Comparison of required and actual grid current for L = 2 mH.

The current supplied to the grid and reference currents are shown in Fig. 9 and 10.

Due to the fact that the instantaneous active power in p-q theory has real physical meaning (product of the same phase voltage and current) there is no error is steady state between required and actual active power to be supplied to the grid (Fig. 7 and 8). However, the imaginary power in the p-q theory has no real physical meaning (product of different phase current and voltage) and by considering it to be the reactive power the error between required and actual imaginary power is generated. This error is more significant as the grid current distortion is increasing.

V. Conclusion

The p-q theory used to power control of grid-connected single-phase inverter is presented. Simulation is used to evaluate the quality of proposed control structure. It is shown that p-q theory can be used for fast power control with acceptable error in reactive power control for small grid current distortion. It is required by grid codes that the current supplied to the grid has small THD. It results in almost sinusoidal grid current. For sinusoidal voltage and current and balanced load the p-q theory can be used as fairly accurate replacement for general power theory. The proposed control structure will be verified by measurement on laboratory model of small grid-tied cascade inverter as simulations show promising results.

Acknowledgment

This work was supported by the Slovak Research and Development Agency under the contract No. APVV-0185-10.

References

[1] Yaosuo Xue, K.C. Divya, G. Griepentrog, M. Liviu, S. Suresh, M. Manjrekar, “Towards Next Generation Photovoltaic Inverters” [Online]. IEEE, 2011

[2] NREL: A Review of PV Inverter Technology Cost and Performance Projections, [Online].

[3] M. Calais, V. G. Agelidis and M Meinhardt, "Multilevel Converters for Single-Phase Grid Connected Photovoltaic Systems: An Overview, [Online].

[4] H. Akagi, Y. Kanzawa, A. Nabae, “Instantaneous Reactive Power Compensators Comprising Switching Devices without Energy Storage Components”, IEEE Transactions on Industry Applications, Vol. IA-20, No. 3, 1984

[5] J.-W. Choi, Y.-K. Kim, H.-G. Kim, “Digital PLL control for single-phase photovoltaic system” [Online]. IEEE, 2006

[6] B. Meersman, J. De Kooning, T. Vandoorn, L. Degroote, B. Renders, L. Vandevelde, “Overview of PLL methods for distributed generation units” [Online]. IEEE, 2010

[7] J. Guo, X. Xiao, S. Tao, “Discussion on Instantaneous Reactive Power Theory and Currents' Physical Component Theory”, [Online]. IEEE, 2012

[8] G. S. Perantzakis, F. H. Xepaps,S. A. Papathanassiou, S. N. Manias, “A Predictive Current Control Technique for Three-Level NPC Voltage Source Inverter“, in Power Electronics Specialists Conference, 2005. PESC '05. IEEE 36th, 2005, pp. 1241 – 1246.

[9] R. Teodorescu, M. Liserre, P. Rodríguez, “Grid Converters for Photovoltaic and Wind Power Systems”, Wiley, ISBN 978-0-470-05751-3, 416 pp, 2011

[10] J. Rodríguez, P. Cortés, “Predictive Control of Power Converters and Electrical Drives”, Wiley, ISBN 978-1-1199-6398-1, 244 pp, 2012

[11] S. Khomfoi, L. M. Tolbert: Multilevel Power Converters, from H.Rashid: Power Electronics Handbook, second release, Elsevier, 2007, ISBN10: 0-12-088479-8, pp. 451 – 482.

Unsupervised Linear Discriminant Subspace Training Based on Heuristic Eigenspectrum Analysis of Speech

Peter VISZLAY (4^{th} year),
Supervisor: Jozef JUHÁR

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

peter.viszlay@tuke.sk

***Abstract*—Linear discriminant analysis (LDA) is a popular linear feature transformation applied in current automatic speech recognition (ASR) to extract robust speech features with good discriminative properties. Since LDA is a supervised method, it requires class labels for the training process. Therefore, the unsupervised training of LDA is still discussed research problem because many issues have not been answered yet. In this paper, we present a novel, unsupervised extension of LDA (ULDA), which enables to train LDA without using any supervision. The proposed method is represented by an unsupervised heuristic algorithm, which assigns class labels to feature vectors. The class definition is based on a data-driven eigenspace analysis of MFCC features. The parameters of LDA are then computed in unsupervised manner. The method is thoroughly evaluated in several scenarios using a Slovak corpus in phoneme-based large vocabulary continuous speech recognition (LVCSR). The method is compared to the state-of-the-art MFCCs and to LDA features. The experimental results show that ULDA features consistently perform better than MFCCs and also slightly improve the recognition accuracy of the conventional LDA.**

***Keywords*—context vector, covariance matrix, discriminant analysis, eigenspectrum, supervision.**

I. Introduction

Linear discriminant analysis (LDA) is a well-known linear transformation method applied in speech recognition [1], [2] to improve the discriminative properties of the baseline features. LDA is a supervised method so appropriate class labels are needed for the training process. The class labels are usually obtained from the automatic time alignment of the continuous speech signal with its corresponding phonetic transcription [3].

In HMM-based recognizers, the speech segments are typically labeled using the forced alignments of the training data computed during the embedded training. This process creates a recognition network for each utterance using the word level transcriptions and the dictionary. The best matching path through the network is found and a lattice containing alignment information is constructed. The lattice is then converted to class label [4]. However, the alignment may not be precise at all conditions because the selected path may not always be the appropriate one. It implies that the LDA class definition based on the forcibly aligned segments can result in vague class labels.

In summary, the unsupervised training of LDA, especially in speech recognition field, is a still open issue because many questions have not been answered yet. There are just several previous works partially related to ULDA. In [5], K-means clustering was used to generate LDA labels for text and gene data. In [6], fuzzy c-means clustering was used to partition samples into categories. Then LDA was used in unsupervised feature extraction (UFLDA). In [7], constrained LDA (CLDA) for classifying hyperspectral images was proposed and extended into an unsupervised version. In [8], k-nearest neighbors were used for ULDA to test handwritten number characters. In other field, unsupervised linear discriminant analysis diameter measurement (ULDM) [9] was proposed to measure retinal vasculatures. In [10], ULDA was used for classification of multi-spectral MRI images. In ASR, there are no relevant work about ULDA, except [11], where K-means clustering was again applied before LDA.

We can conclude that there is still missing a particular methodology in ASR, which enables to get the LDA classes using only the training data without any classical clustering. This fact motivated us to analyse the ULDA problem in-depth. We have built upon our previous works [12], [13] and we proposed an unsupervised extension of LDA. In this paper, we introduce an original algorithm to generate class labels established on rigorous statistical similarities of feature vectors. The proposed method consists of three learning stages. The first one is the unsupervised heuristic algorithm used to create and assign a unique class label to each feature vector. The class definition is based on a data-driven standardized eigenspectrum analysis of MFCC features. The second stage is focused to refinement of the created classes by applying two class-dependent transformations. Finally, at the third learning stage, the parameters of discriminant analysis are computed.

We evaluated the ULDA in phoneme-based large vocabulary continuous speech recognition task. The experiments were conducted using a Slovak corpus. The results are compared to reference MFCC and also to LDA features. It is shown that the features obtained with the new method outperform both the reference MFCC and LDA estimated features.

The rest of the paper is organized as follows. In Section 2, the conventional LDA training is reviewed. Section 3 presents the proposed ULDA method in detail. The experimental setup is given in Section 4. The experimental evaluation of ULDA with results is presented in Section 5. Finally, we state our conclusions including our future intentions in Section 6.

II. Conventional LDA training

According to [1], [2] and [14], the classical LDA estimation can be defined as follows. Suppose a training data matrix $X \in \Re^{N\times n}$ with n column vectors $\boldsymbol{x}_i$, where $1 \leq i \leq n$. The aim of LDA is to find a linear transformation represented by transformation matrix $W \in \Re^{N\times p}$ that maps each vector $\boldsymbol{x}_i$ to a vector $\boldsymbol{y}_i$ in the p-dimensional space. Thus, for W applies $\boldsymbol{x}_i \in \Re^N \rightarrow \boldsymbol{y}_i \in \Re^p;\ \boldsymbol{y}_i = W^T\boldsymbol{x}_i;\ p < N$. Consider that the training data are partitioned into k classes as $X = \{\Pi_1, \ldots, \Pi_k\}$, where the class Π_i contains n_i feature vectors. Notice that $n = \sum_{i=1}^{k} n_i$. The classes can be represented by *class mean vectors* $\boldsymbol{\mu}_i = \frac{1}{n_i}\sum_{\boldsymbol{x}\in\Pi_i}\boldsymbol{x}$ and their *class covariance matrices* $\Sigma_i = \frac{1}{n_i-1}\sum_{\boldsymbol{x}\in\Pi_i}(\boldsymbol{x}-\boldsymbol{\mu}_i)(\boldsymbol{x}-\boldsymbol{\mu}_i)^T$, $i \in \langle 1; k\rangle$, which are defined to quantify the quality of the cluster. Since LDA is applied in class-independent manner, we define the *within-class covariance matrix* Σ_W as the weighted sum of all class covariance matrices:

$$\Sigma_W = n_i \sum_{i=1}^{k}\left[\frac{1}{n_i-1}\sum_{\boldsymbol{x}\in\Pi_i}(\boldsymbol{x}-\boldsymbol{\mu}_i)(\boldsymbol{x}-\boldsymbol{\mu}_i)^T\right]. \quad (1)$$

The *between-class covariance matrix* is used to quantify the covariance between classes. It is defined as follows:

$$\Sigma_B = \frac{1}{n-1}\sum_{i=1}^{k}(\boldsymbol{\mu}_i-\boldsymbol{\mu})(\boldsymbol{\mu}_i-\boldsymbol{\mu})^T, \quad (2)$$

where $\boldsymbol{\mu} = \frac{1}{n}\sum_{i=1}^{k}\sum_{\boldsymbol{x}\in\Pi_i}\boldsymbol{x}$ is the *global mean vector* (computed disregarding the class label information). In speech recognition, $\boldsymbol{x}$ is a *supervector* created by concatenating of J acoustic vectors computed on successive speech frames. The feature vector $\boldsymbol{x}_j$ at the j-th position is spliced together with $\frac{J-1}{2}$ vectors on the left and right as:

$$\mathbf{x} = \left[\ \mathbf{x}[j-\tfrac{J-1}{2}]\ \ \ldots\ \ \mathbf{x}[j]\ \ \ldots\ \ \mathbf{x}[j+\tfrac{J-1}{2}]\ \right], \quad (3)$$

where J is typically $3 \approx 11$ frames. The covariance matrices are used to formulate the optimization criterion of LDA:

$$J(W) = \frac{|W^T\Sigma_B W|}{|W^T\Sigma_W W|}. \quad (4)$$

This optimization is equivalent to the eigenvalue problem $\Sigma_B\boldsymbol{v} = \lambda\Sigma_W\boldsymbol{v}$, for $\lambda \neq 0$, where $\boldsymbol{v}$ is a matrix of eigenvectors and λ represents the set of eigenvalues. In practice, the solution can be obtained by applying an eigendecomposition to the matrix $\Sigma_{WB} = \Sigma_W^{-1}\Sigma_B$. Finally, the matrix of eigenvectors $\boldsymbol{v}$ is used to create the LDA matrix.

III. Proposed method - Unsupervised linear discriminant analysis of speech

In this chapter we describe the proposed method in detail. *The main objective of the algorithm is to create and assign a unique label to each training feature vector without using any supervision*. This heuristic process uses to define the classes only the training corpus and it is applied for all feature vectors in the same way repeatedly.

A. Class definition in ULDA

The class definition in ULDA is based on the fundamental assumption that each N-dimensional feature vector $\boldsymbol{x} = [x_1, \ldots, x_N]$ can be geometrically interpreted as an N-point data cluster in one dimensional space. The scatter σ_x^2 of the cluster explained along the coordinate axis x can be exactly quantified as $\sigma_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i-\bar{x})^2$, where $\bar{x} = \frac{1}{N}\sum_{i=1}^{N}x_i$ is the mean. The vector $\boldsymbol{x}$ can be rearranged to feature matrix X with row and column dimensions $N_r = 2$ and N_c, respectively ($N_r \leq N_c$, where $N_c = \frac{N}{N_r}$). Then applies that $\forall x_i \in \boldsymbol{x} \mapsto x_{i,j} \in X,\ i \in \langle 1; N_r\rangle$, $j \in \langle 1; N_c\rangle$. Then X contains N_c elementary column feature vectors $X = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{N_c}]$. In the two dimensional space xy, two scatters σ_x^2 and σ_y^2 can be analysed. The *total scatter* Σ^2 of the cluster can be defined as $\Sigma^2 = \sigma_x^2+\sigma_y^2$. It means that σ_x^2 and σ_y^2 have in the whole scatter amount a certain significance. Therefore, we define the parameter called *scatter significance*. It can be denoted for σ_x^2 as $\Delta(\sigma_x^2)$ and for σ_y^2 as $\Delta(\sigma_y^2)$. They are defined as:

$$\Delta(\sigma_x^2) = \frac{\sigma_x^2}{\sigma_x^2+\sigma_y^2} = \frac{\sigma_x^2}{\Sigma^2}, \quad (5)$$

$$\Delta(\sigma_y^2) = \frac{\sigma_y^2}{\sigma_x^2+\sigma_y^2} = \frac{\sigma_y^2}{\Sigma^2}. \quad (6)$$

The parameter $\Delta(\sigma_x^2)$ represents the amount of scatter explained along the axis x and analogously, $\Delta(\sigma_y^2)$ represents the amount of scatter explained along the axis y. The values of $\Delta(\sigma_x^2)$ or $\Delta(\sigma_y^2)$ lie in the interval $(0; 1)$ and applies that $\Delta(\sigma_x^2) + \Delta(\sigma_y^2) = 1$. For better explanation, the percentage measure of $\Delta(\sigma_x^2)$ and $\Delta(\sigma_y^2)$ can be defined as $\Delta(\sigma_x^2)_{[\%]} = \Delta(\sigma_x^2) \times 100\%$ and $\Delta(\sigma_y^2)_{[\%]} = \Delta(\sigma_y^2) \times 100\%$. The maximum significance $\Delta_{max}(\sigma^2)$ can be found as $\Delta_{max}(\sigma^2) = \max\left(\Delta(\sigma_x^2), \Delta(\sigma_y^2)\right)$. But there is a problem related to the fact that $\Delta_{max}(\sigma^2)$ may appertain to other coordinate axis for each feature vector. This means that for one vector $\Delta(\sigma_x^2)$ corresponds to $\Delta_{max}(\sigma^2)$ and for the next vector $\Delta(\sigma_y^2)$ may correspond to $\Delta_{max}(\sigma^2)$. Therefore, the sorting procedure based on the parameter $\Delta_{max}(\sigma^2)$ would be inconsistent one.

We found that application of principal component analysis (PCA) [15] to the feature matrix X in order to standardize the original scatter space leads to consistent sorting scheme. PCA is applied to X as follows. The covariance matrix of X with removed mean is computed as $C = \frac{1}{N_c-1}\sum_{i=1}^{N_c}(\boldsymbol{x}_i-\mu)(\boldsymbol{x}_i-\mu)^T$, where $\mu = \frac{1}{N_c}\sum_{i=1}^{N_c}\boldsymbol{x}_i$. An eigendecomposition is then applied to covariance matrix, which results in a set of eigenvectors $\boldsymbol{u} = \{\boldsymbol{u}_1, \boldsymbol{u}_2\}$ and eigenvalues $\lambda = \{\lambda_1, \lambda_2\}$. Then applies that $C\boldsymbol{u}_i = \lambda_i\boldsymbol{u}_i$, $i \in \langle 1; 2\rangle$. According to PCA, the eigenvalues are sorted in descending order as $\lambda_1 > \lambda_2$. The i-th eigenvalue represents the variance explained along the i-th principal component (PC) in PCA space [15]. Based on this fact, the total variance (scatter) Λ in PCA space can be defined as $\Lambda = \lambda_1 + \lambda_2$. It is clear that the total scatter (Σ^2) in the original space and the total scatter in PCA space (Λ) must be equal ($\Sigma^2 = \Lambda$) because the scatter of the cluster must be preserved in PCA space. It can be also showed that $\Lambda = \text{trace}(C)$. In PCA space, the following significance can be defined for each eigenvalue as:

$$\Delta(\lambda_i) = \frac{\lambda_i}{\lambda_1+\lambda_2} = \frac{\lambda_i}{\Lambda}, \quad (7)$$

or in percentage $\Delta(\lambda_i)_{[\%]} = \Delta(\lambda_i) \times 100\%$. It is known that the set of eigenvalues is called as eigenspectrum [15]. Therefore, we called the $\Delta(\lambda_i)$ parameter as *spectral significance*. It represents the amount of variance explained by i-th principal component (PC) in PCA space. Since $\lambda_1 > \lambda_2 \Rightarrow \lambda_{max} = \lambda_1$. It is clear that the maximum spectral significance $\Delta_{max}(\lambda)$ always corresponds to the first eigenvalue, thus $\Delta_{max}(\lambda) =$

$\Delta(\lambda_1)$. This means that $\Delta_{max}(\lambda)$ is always related to 1. PC *for each feature vector* so the dominance of $\Delta_{max}(\lambda)$ is invariant in the PCA space. In comparison with the previous scatter analysis in the original space, the analysis based on PCA is a consistent scheme to determine the scatter properties of the data. The scatter relations in the feature matrix in the original and also in the PCA space are geometrically interpreted in the Fig. 1 for a general example.

The unsupervised sorting of training vectors in ULDA is based on the idea that *feature vectors with equal or very similar maximum spectral significance can be grouped together in separate classes*. The created classes are considered as ULDA classes usable to estimate the parameters of discriminant analysis. The ULDA classes fundamentally differ from the conventional LDA classes defined on phone or triphone level. This classes are created according to heuristic analysis of the statistical properties of the feature vectors. Therefore, we called the ULDA classes as *statistical classes*.

B. Integration to speech recognition framework

Each training MFCC vector $\boldsymbol{x}_i$, $i \in \langle 1; n\rangle$ with dimension $N = 13$ is extended with zero coefficient to dimension $N_x = 14$ in order to rearrange it to feature matrix X_i with dimension 2×7. In the next step, PCA is applied to X_i according to Section III-A. Then, the maximum spectral significance $\Delta_{max}(\lambda)$ is computed and converted to percentage value $\Delta_{max}(\lambda)_{[\%]}$ according to (7). In order to create a class label `L` for the current feature vector $\boldsymbol{x}_i$, the value of $\Delta_{max}(\lambda)_{[\%]}$ is rounded down to the nearest integer as $\texttt{L} = \texttt{floor}(\Delta_{max}(\lambda)_{[\%]})$. For example, if $\Delta_{max}(\lambda)_{[\%]} = 65.75\%$, then `L=65`.

In order to take into account the contextual information (as it is typical in LDA) we have done several experiments with classical supervectors composed according to (3). The results showed that supervectors are not suitable for ULDA due to high dimension of the classes (≈ 100). But we found that the *right context* (Rc) of the current vector is a suitable one. We evaluated different lengths of right context, see Section V.

The heuristic sorting procedure for each feature vector (with its right context) is performed repeatedly, which results in several statistical classes $\widehat{\Pi}_i$, $i \in \langle 1; k'\rangle$. Each class is represented by class matrix $A(\widehat{\Pi}_i)$. The number of created classes is $k = 50$ because $\Delta_{max}(\lambda)_{[\%]} + \Delta(\lambda_2)_{[\%]} = 100$. This means that the labels are defined from `L=50` to `L=99`. It can be concluded that *the number of statistical classes in ULDA is comparable with the number of phonetic LDA classes* ($k \approx k' \Rightarrow 45 \approx 50$). In addition, *we experimentally found that the maximum spectral significance is always larger than the maximum scatter significance for each feature vector*, thus $\Delta_{max}(\lambda) \geq \Delta_{max}(\sigma^2)$.

The created classes can be directly used to estimate the ULDA parameters. We found that the class quality can be improved by application of class-dependent PCA (CD-PCA) transformation without dimension reduction, which decorrelates the feature vectors in the class and improve the statistical properties of the class covariance matrix. CD-PCA is applied separately to each class as $A(\widehat{\Pi}_i)_{PCA} = \Theta_i^T A(\widehat{\Pi}_i)$, where Θ_i is the full-rank PCA matrix estimated from the i-th class. We also found that PCA classes can be further refined by class-dependent whitening (CD-W) as $A(\widehat{\Pi}_i)_{white} = \Omega_i(PCA)A(\widehat{\Pi}_i)_{PCA}$, where $\Omega_i(PCA)$ is the i-th whitening matrix estimated from the i-th PCA class.

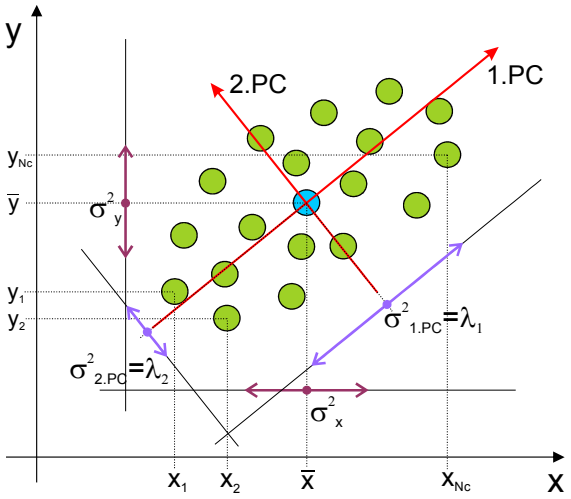


Fig. 1. The scatter relationships between the original and PCA 2D space

The class refinement is followed by ULDA estimation described in Section II, which results in ULDA transformation matrix W. This matrix is used to transform the train and test set independently, without any context vectors (dimension 13×13). The right context is used only at the ULDA training.

IV. Experimental Setup

The Slovak corpus [16] used in our study contains approx. 100 hours ($\approx$ 20 million frames) of spontaneous Slovak parliamentary speech. About 37000 utterances from 120 speakers were used for training and another 880 ones were used to test. The phone segmentation of 45 Slovak phones from embedded training and automatic phone alignment was obtained.

The speech signal was preemphasized and windowed every 10ms using Hamming window of length 25ms. Fast Fourier transform was applied to the windowed segments. Melfilterbank analysis with 26 channels were applied. Then, 39-dimensional MFCC vectors were used to baseline acoustic modeling. For LDA-based processing, 13-dimensional MFCC vectors were used. After transformation, 13 LDA features were retained and expanded with delta and acceleration coefficients. The LDA classes corresponded to 45 Slovak phonemes.

Our recognition system used context independent monophones modeled using a three-state left-to-right Hidden Markov Models (HMMs). The number of Gaussian mixtures per state was a power of 2, starting from 1 to 256. The number of trained monophone models corresponds to the number of phonemes. In order to test the acoustic models a bigram language model was built. The vocabulary size was 125k words. The accuracies in the evaluation process were computed as the ratio of number of all word matches to number of reference words. The acoustic model training and testing using HTK Toolkit [4] was performed.

V. Experimental Evaluation and Results

In this section, the proposed ULDA method is experimentally evaluated. The maximum recognition accuracies achieved by ULDA are listed in Table I. These accuracies are graphically evaluated in the Fig. 2. From this table and figure it can be seen that ULDA estimated from base classes without any further processing achieves the lowest performance. It is clearly demonstrated that CD-PCA consistently improves the

TABLE I
RECOGNITION ACCURACIES OF ULDA

Num. of mixtures	Base classes	Right context	PCA classes	Right context	White classes	Right context	Max. ULDA
1	80.41%	Rc=4	82.45%	Rc=1	82.60%	Rc=2	82.60%
2	81.75%	Rc=1	84.35%	Rc=0	84.72%	Rc=2	84.72%
4	84.51%	Rc=1	86.96%	Rc=0	87.37%	Rc=0	87.37%
8	86.53%	Rc=2	89.34%	Rc=4	89.44%	Rc=0	89.44%
16	87.87%	Rc=0	90.72%	Rc=4	90.49%	Rc=0	90.72%
32	88.84%	Rc=2	91.39%	Rc=4	91.31%	Rc=2	91.39%
64	90.00%	Rc=4	92.10%	Rc=2	92.10%	Rc=3	92.10%
128	90.75%	Rc=0	92.46%	Rc=4	92.54%	Rc=0	92.54%
256	91.43%	Rc=0	92.70%	Rc=4	92.92%	Rc=0	92.92%

TABLE II
COMPARISON OF ULDA TO REFERENCE MFCCS AND LDA

Num. of mixtures	Ref. MFCC	Ref. LDA	Abs. dif.	Max. ULDA	Dif. MFCC	Dif. LDA
1	82.32%	81.37%	−0.95%	82.60%	**+0.28%**	**+1.23%**
2	83.26%	83.60%	+0.34%	84.72%	**+1.46%**	**+1.12%**
4	85.06%	87.11%	+2.05%	87.37%	**+2.31%**	**+0.26%**
8	87.77%	88.47%	+0.70%	89.44%	**+1.67%**	**+0.97%**
16	89.53%	90.03%	+0.50%	90.72%	**+1.19%**	**+0.69%**
32	90.83%	90.88%	+0.05%	91.39%	**+0.56%**	**+0.51%**
64	91.48%	91.80%	+0.32%	92.10%	**+0.62%**	**+0.30%**
128	92.37%	92.48%	+0.11%	92.54%	**+0.17%**	**+0.06%**
256	92.50%	92.90%	+0.40%	92.92%	**+0.42%**	**+0.02%**

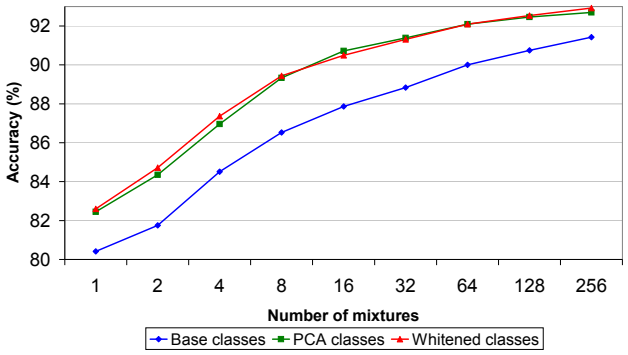


Fig. 2. Influence of the class quality refinement to the performance of ULDA

ULDA performance for all mixtures. The maximum improvement achieved by CD-PCA is +2.85% for 16 mixtures. It is also proven that class-dependent whitening applied to PCA classes leads to improvement (max. +0.41% for 4 mixtures).

Table II presents the main comparison of ULDA with MFCCs and conventional LDA. Note that LDA was estimated with context $J = 3$ and all models were 39-dimensional. As it was expected, LDA improves the baseline MFCC performance. The results show that ULDA features significantly outperform the state-of-the-art MFCCs for all mixtures (max. +2.31%). But what is more important, ULDA also effectively outperformed LDA for all mixtures. The maximum improvement of LDA by ULDA is +1.12% for 2 mixtures. From the final results it can be observed that the improvement of LDA decreases with increasing number of mixtures. The results from the table are graphically compared in Fig. 3.

VI. CONCLUSION AND FUTURE WORK

In this paper we presented a novel unsupervised extension of LDA applied in ASR, which does not require the traditional training class labels. We showed that our proposed methodology completely performs better than LDA. In the near future

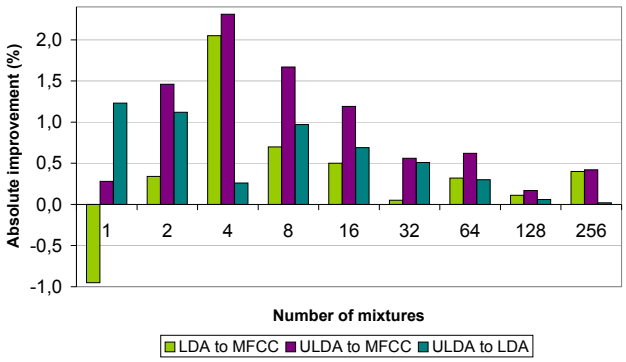


Fig. 3. Comparison of absolute improvements achieved by ULDA

we want to further refine the ULDA algorithm and we want to try the eigenspace analysis in higher dimensional PCA space.

ACKNOWLEDGMENT

This work has been performed under research project ITMS-26220220155 supported by the Research & Development Operational Programme funded by the ERDF.

REFERENCES

[1] R. Haeb-Umbach and H. Ney, "Linear discriminant analysis for improved large vocabulary continuous speech recognition," in *Proc. of the IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing, ICASSP'92*, San Francisco, CA, 1992, pp. 13–16.

[2] J. Pylkkönen, "LDA based feature estimation methods for LVCSR," in *Proc. of the 9th Intl. Conf. on Spoken Language Processing, INTERSPEECH'06*, Pittsburgh, PA, USA, 2006, pp. 389–392.

[3] H. Leung and V. W. Zue, "A procedure for automatic alignment of phonetic transcriptions with continuous speech," in *Proc. of the Intl. Conf. on Acoustics, Speech, and Signal Processing, ICASSP'84*, 1984, pp. 73–76.

[4] S. Young, G. Evermann, M. Gales, T. Hain, D. Kershaw, X. A. Liu, G. Moore, J. Odell, D. Ollason, D. Povey, V. Valtchev, and P. Woodland, *The HTK Book (for HTK Version 3.4)*, Dec. 2006.

[5] C. Ding and T. Li, "Adaptive dimension reduction using discriminant analysis and K-means clustering," in *Proc. of the 24th Intl. Conf. on Machine Learning, ICML'07*, New York, USA, 2007, pp. 521–528.

[6] C. Li, B.-C. Kuo, and C.-T. Lin, "LDA-Based Clustering Algorithm and Its Application to an Unsupervised Feature Extraction," *IEEE Transactions on Fuzzy Systems*, vol. 19, no. 1, pp. 152–163, 2011.

[7] Q. Du, "Unsupervised real-time constrained linear discriminant analysis to hyperspectral image classification," *Pattern Recognition*, vol. 40, no. 5, pp. 1510–1519, 2007.

[8] H. Tang, T. Fang, P. Shi, and G. Tang, "Unsupervised linear discriminant analysis," *Journal of Shanghai Jiaotong University (Science)*, vol. 11, no. 1, pp. 40–42, 2006.

[9] D. K. Kumar, B. Aliahmad, and H. Hao, "Retinal vessel diameter measurement using unsupervised linear discriminant analysis," *ISRN Ophthalmology*, vol. 2012, pp. 1–7, 2012.

[10] G.-C. Lin, W.-J. Wang, and S. Y. Sun, "Automated classification of multi-spectral MR images using Linear Discriminant Analysis," *Computerized Medical Imaging and Graphics*, vol. 34, pp. 251–268, 2010.

[11] T. Giannakopoulos and S. Petridis, "Unsupervised speaker clustering in a linear discriminant subspace," in *Proc. of the 9th Intl. Conf. on on Machine Learning and Applications, ICMLA 2010*, Washington DC, USA, 2010, pp. 1005–1009.

[12] P. Viszlay, J. Juhár, and M. Pleva, "Alternative phonetic class definition in linear discriminant analysis of speech," in *Proc. of the 19th Intl. Conf. on Systems, Signals and Image Processing (IWSSIP)*, 2012, pp. 655–658.

[13] J. Juhár and P. Viszlay, *Modern Speech Recognition Approaches with Case Studies*. InTech, 2012, ch. Linear Feature Transformations in Slovak Phoneme-Based Continuous Speech Recognition, pp. 131–154.

[14] N. Kumar, "Investigation of silicon auditory models and generalization of linear discriminant analysis for improved speech recognition," Ph.D. dissertation, Johns Hopkins Universtiy, Baltimore, Maryland, 1997.

[15] I. T. Jolliffe, *Principal Component Analysis*. New York, USA: Springer-Verlag, 1986.

[16] S. Darjaa, M. Cerňak, Š. Beňuš, M. Rusko, R. Sabo, and M. Trnka, *Rule-based triphone mapping for acoustic modeling in automatic speech recognition*, ser. LNCS, 2011, vol. 6836 LNAI.

User interface of intra-abdominal pressure measuring system

[1]*Michal JURČIŠIN (4th year)*
Supervisor: [2]Stanislav SLOSARČÍK

[1, 2] Dept. of Technologies in Electronics, FEI TU of Košice, Slovak Republic

[1]michal.jurcisin@tuke.sk, [2]stanislav.slosarcik@tuke.sk

***Abstract* — The article deals with process of development of automation system for non-invasive measurement of intra-abdominal pressure using designed software. Designed software tool was used for defining of optimal pressure sensor parameters which is implemented into measurement system. Pressure sensor consists of glass tube with evaporated thin film electrodes. Developed software tool enables evaluation of measurement system features and parameters, which was tested in laboratory conditions. In the paper the software tool is presented which is able to control the peripherals of measurement system manually as well as automatically with possibility of measured data archiving.**

***Keywords* — capacitive sensor, intra-abdominal pressure, automation control, pressure sensor.**

I. Introduction

Abdominal compartment syndrome (AbCS) is caused by the increase of intra-abdominal pressure (IAP) from tissue tumescence or from accumulation of free liquid in abdominal cavity. When it is not timely diagnosed and it is not cured, it may result in multi-organ collapse and death [1]. For measuring of IAP exist direct and indirect methods. Indirect method of IAP measurement is based on Foley catheter inserted inside urinary bladder. This method is based on the fact that urinary bladder works as a passive transmitter IAP on internal water filling. Empty urinary bladder is filled with 20-50 ml of sterile saline via catheter. [2]

II. Concept of Automation System for Intra-abdominal Pressure Measurement

For using of developed automation measurement system in clinical praxis it is necessary to state the basic requirements.

Due to application field it is necessary to ensure sterility of the system during measurement. Distribution of intravesical pressure grades used in clinical practice is listed in Table I. In the second grade i.e. at a pressure of 2.66 kPa it is necessary to decompress the abdominal cavity what is reached by cutting the abdominal wall and keep it opened. Measurement of IAP is carried out at one hour intervals with the following procedure:

- system preparation for measurement - prefilling of bladder and system tubes with exact volume of saline,
- realization of measurement in the time interval of several minutes,
- drying of system.

TABLE I
Distribution of pressure grades used in clinical practice

Grade	cmH_2O	p [kPa]
1.	16 – 20	1.59 – 1.99
2.	21 – 27	2.13 – 2.66
3.	28 – 34	2.79 – 3.33
4.	> 35	> 3.33

III. Capacitive Pressure Sensor

For realization of pressure sensor was used thin film technology. Capacitive pressure sensors were realized by glass tubes with 4-9 mm inner diameter. On the outer surface of the glass tube the two thin-film electrodes was deposited by vapor deposition. The copper was used as a material for deposition. The shape of the electrodes is shown in Fig. 1. As a dielectric of capacitive sensor the saline was used (0.9% NaCl aqueous solution).

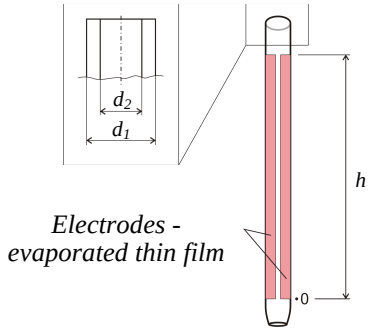


Fig. 1. Capacitive pressure sensor (d_1 – outer tube diameter, d_2 – inner tube diameter, h – height of evaporated electrode

Parameters of used glass tubes were analyzed in [4]. Measured capacity varied depending on the tube diameter in the range of 11.7 pF - 25 pF without dielectric corresponding to the 0 kPa pressure, and 14.3 - 132 pF with a dielectric at 300 mm height corresponding to the 2.94 kPa pressure according to [4].

IV. Description of Automation System for Measurement of Intra-Abdominal Pressure

Block diagram of developed automation system for measuring of IAP by noninvasive method is shown in Fig. 2.

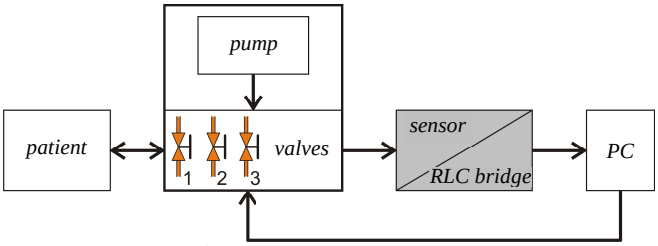


Fig. 2. Block diagram of automation measurement system.

In time interval between measurements emptying of system, as well as patients bladder occur. The first step of measurement process is the filling of tubes between sensor and pump with defined volume of saline. In the next step is filled patients bladder with 30 ml of saline. Measurement starts after filling. Pre-filled tubes ensure continuity of liquid column and the urine interference on sensor is eliminated. Described procedure represents one measurement cycle. The valves consist of servomotor which can be logically controlled. This solution was realized in order to ensure the sterility of the measuring system.

Operation of valves and infusion pump is provided by control electronics on the basis of program. Control electronics also provide communication with monitoring software in computer.

V. Monitoring Software for Automation Measurement System

For simple control of automation measurement system was developed monitoring software. Monitoring software provide graphical representation of measured pressure respectively sensor capacitance. Other features of created application are the archiving of measured results with possibility of data re-viewing and control of measurement system peripherals. Graphic user interface of monitoring software is shown on Fig. 3.

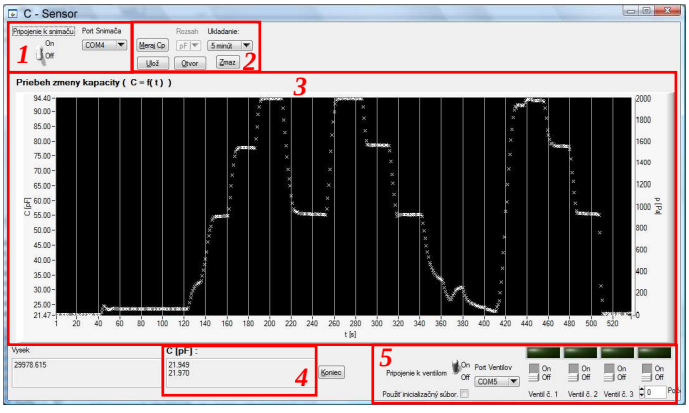


Fig. 3. Graphic user interface of software for monitoring and control of automation system for measurement of intra-abdominal pressure.

User interface of monitoring software consists of following parts (Fig. 3):

1 – connection to capacitance measuring instrument – before connection to capacitance measuring instrument it is possible to make necessary settings (port number),

2 – saving and loading of archived files – possibility of saving of measured data, loading of archived files,

3 – graphical representation of measured data – displaying of measured capacity (left Y axis), resp. pressure (righty Y axis) in time dependence (X axis). Time axis is dynamically adjusted.

4 – list of measured data in text format – at this part is displayed text list of measured results. The list is updated after every measurement.

5 – connection settings and buttons for control of valves and pump of automation measurement system – enables the connection to peripheral parts of system, peripherals can be controlled manually or automatically.

A. *Archiving of measured data*

Created monitoring software provides manual and automatic archiving of measured data. After elapse of preset time period the automatic data saving occurs. Other possibility to data archiving is manual saving. Saved files can be edited as needed. Structure of saved file is shown on Fig. 4.

```
11:00:32                14. 12. 2012

t [s]    Cp [pF]

1        59.605
2        59.625
3        59.585
4        59.603
5        59.588
6        59.619
7        59.614
8        59.556
9        59.547
10       59.622
```

Fig. 4. Example of saved file.

First row of file contains information about time and date of archiving. Next item is header of table. First column of table is number of seconds from the beginning of measurement. Second column contains capacity value for corresponding time.

B. *Control of automation measurement system peripherals*

Peripherals of measurement system can be controlled manually or automatically. Before connection of monitoring software to the peripherals it is necessary to set manual or fully automatic control. Indicator of automatic control is check box *„Use INI file“*. In case of manual control the user can directly control peripherals using buttons according figure 5.

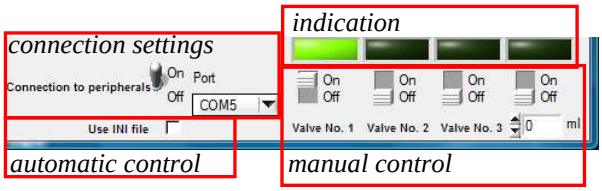


Fig. 5. Controls for handling of peripherals.

Peripherals can be controlled by four switches. First three switches control valves. Fourth switch control the infusion pump. Above each switch is located indicator which determines the peripheral's state. In case of automatic control it is not possible to change the switches state resp. peripherals state by pressing it. Automatic control of peripherals uses initialization file. Initialization file structure is shown on Fig. 6.

```
t       V1    V2    V3    PUMP
[m:s]
00:01   0     0     0     0
00:02   0     0     1     0
00:03   0     0     1     5
00:15   1     1     0     0
00:16   1     1     0     10
00:45   1     1     0     0
00:46   1     1     1     0
00:47   1     0     1     27
01:57   1     0     0     0
09:59   0     0     0     0
19:59   0     0     0     0
EOF
```

Fig. 6. Structure of initialization file for automatic control of system peripherals.

The file consists of five columns. First column represents the time since beginning of actual measurement cycle. At the time defined by first column, peripherals state is changed according to state of other columns. Columns 2, 3 and 4 define states of valves. Column number five defines volume (in ml) of saline which has to be injected into the system. Initialization file shown on Fig. 6 presets 20 minutes measurement cycle. During first 2 minutes system initialization is realized. In next 8 minutes is realized measurement. Subsequently system is drained until the end of the measurement cycle. After completion of the measurement cycle is procedure automatically started from the beginning of initialization file.

VI. Conclusion

Measurement of IAP by presented measuring system contributes to workload reduction of medical staff thanks to fully automatic design. Created software enabled the development of suitable capacitive sensor, which was implemented into measurement system. Developed software provide user friendly interface between measured data and data comprehensible for medical staff. This feature enables unification and simplification of measurement process. In addition the presented developed software tool provides several methods for data archiving as well as manual and fully automatic control of system peripherals. (Project VEGA 1/0059/12)

References

[1] T. Molčányi, J. Magdo, R. Raši, T. Cestická, J. Michlík, A. Molčányiová, "Možnosti ošetrovania laparostomie při abdominálnom kompartment syndrome" in *Novinky v anasteziológii, intenzívnej medicine a algeziológii*, Prešov, Náuka, 2005, s. 122-127, ISBN 80-89038-37-9.

[2] F. Obeid et al., "Increases in intra-abdominal pressure affect pulmonary compliance", Arch Surg, 1995. 130(5): p. 544-7; discussion 547-8.

[3] I.L. Kron, P.K. Harman, S.P. Nolan, "The measurement of intra-abdominal pressure as a criterion for abdominal re-exploration".

[4] M. Jurčišin, S. Slosarčík, P. Cabúk, "Automation measurement of intra-abdominal pressure", In: Advanced Numerical Modelling: 28. - 31.8.2011, Zielona Góra. – Warsaw: Electrotechnical Institute, 2011, p. 95-96., ISBN 978-83-61956-02-0

Section:
Electrical & Electronics Engineering
Poster form

A concept of the remote control of the laboratory equipment and software tools

[1]*Dominik DEMETER (3rd year)*
Supervisor: [2]Juraj BANSKÝ

[1,2]Dept. of Technologies in Electronics, FEI TU of Košice, Slovak Republic

[1]dominik.demeter@tuke.sk, [2]juraj.bansky@tuke.sk

***Abstract*—This paper is concerning with a concept of a system, which is intended for controlling access to several software tools and laboratory equipment. The timeframe, when the equipment and software tools are available for students during the class is very short and is enough only for get familiar with the user interface of the program, but is very short for serious work. This problem is multiplied by the fact, that some software tool is available only in limited number of licenses due to the high price of it. Sometimes a usable demo version of the tool is missing, too. In this case, it is no space at the class for real work with these tools, what negatively affects the quality of education and is not corresponding with the mission of the University to well prepare students for their future work.**

***Keywords*— e-learning, remote control, remote desktop access, reservation system, virtual laboratory.**

I. INTRODUCTION

University must produce graduates, who are prepared for working with different tools to have enough practical and theoretical knowledge. The faculty and its departments are spending a huge amount of money to modernize the laboratory equipment and software tools, which are used for research, as well as for education purposes. Theoretically, the students of the faculty have the right to legally access and use the laboratory equipment and software tools for education needs.

In practice, the students have no chance to use their right, because in the education process they have only limited time, when they can use these tools. Sometimes it is only limited to class during the lessons. This fact negatively affects the quality of thesis, as well as the quality of the graduates.

To improve the quality of education we need to find the optimal way, how to connect the laboratory equipment to the computer network and how to control them remotely. We need to do this also for software tools. Increasing the time, when the students can work with real software tools will increase his skill to working with it.

II. CONNECTING LABORATORY EQUIPMENT TO THE INTERNET

A. Communication interfaces of laboratory equipment

Laboratory equipment can be controlled remotely through computer network in several ways. Today, most of the manufacturers provide their laboratory equipment with a number of different communication interfaces. The most used communication interface is the serial communication, which is represented by RS-232, or the newer USB port. Some laboratory equipment, especially the measurement tools, has implemented ethernet interface, which allows controlling them remotely through internet. But this option is not a standard and a lot of manufacturers offer this option only for a relatively high price and sometimes, the buyer forgets to order this extra. Because this fact the number of equipment with built-in remote access option is limited and increases slowly.

The laboratory equipment with integrated serial communication interface is intended to use with a personal computer connected through this communication interface or at least, to read out some measured data. The manufacturer provides a controlling application, which allows controlling of most of the function of the laboratory equipment and remote reading the data from it. The disadvantage of this type of communication is the short distance between the equipment and the connected computer. The operator must be near the equipment to read some data or to control the experiment. On the other side, it is useful to have at least this type of communication ability, because it helps the processing of measured data, logging the experiment, etc.

B. Ethernet connection

The case when the laboratory equipment has a built-in Ethernet interface is ideal, because it indicates, that the manufacturer has calculated with the option of remote control, and the whole equipment is optimized for this type of usage. The manufacturer also solved the problem of controlling the device by user interface.

These types of equipment can be controlled trough built-in web interface or trough some software application installed on the client's computer.

More difficult is the situation, when the equipment has only some local communication interfaces. On the market, there exist a several number of products, which can translate the serial communication to network communication. This allows emulating a serial communication port on the client's computer, which acts as a serial port, which is connected directly to the laboratory equipment. The translation between the serial and ethernet communication is done by a com port emulator on the side of the client's computer and by the

hardware on the side of the laboratory equipment (see Fig. 1).

By combination of this type of software and hardware, it is

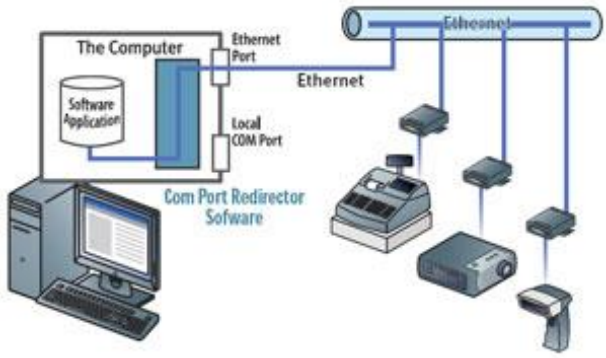


Fig. 1. Example of use the Lantronix XPort and ComPort Redirector software for accessing and remote controlling the laboratory equipment with only serial communication interface through Ethernet network [1].

possible to connect the equipment through ethernet or internet to a computer without distance limits. The attached equipment acts like as it was connected directly to the computer.

A big issue with local communication interfaces, such as RS-232, USB, etc., is, that in most cases this communication interfaces can be accessed only by the software, which was bundled with the equipment. Sometimes it is third party software customized for the equipment manufacturer and is protected by the law and can't be freely copied and installed on multiple computers. In most cases, the usage of this software is protected by a hardware USB key, which must be plugged in the computer, on which the software should run.

III. Remote control of laboratory equipment

Connecting the laboratory equipment to computer network is only the first step on the way to remote controlling of the equipment. It is also important to offer the users an easy to use user interface, which enables the most of functions of the equipment.

It is nearly impossible to create a universal user interface, which will allow controlling of several types of equipment from different manufacturer. It is also not effective to develop individual user interface for the equipment, apart. For the users, the best solution is to use the software, which was bundled with the equipment. But, as I mentioned earlier, this software is mostly not accessible freely and cannot running on several computers at the same time.

The best solution appears to allow users remotely access to the desktop of the computers, which are directly connected to the laboratory equipment (trough RS-232 or USB). On these computers, the bundled software can be installed without any limitations, including the attached USB hardware key. This computer can be locked in the laboratory and nobody must care about the security of the hardware key or the limitations of the license agreement about creating of copies. This kind of access provides the option to control all functions of the equipment.

From the end user's point of view this solution is the most user's friendly, because they can work with the standard software tool and they do not need to learn to use another application for the same thing. By this way, the students can obtain experience with the tools, which they can meet in their future work.

IV. Increasing the effectivity of the software usage

Laboratory software and CAD tools is able to save time and money, if it is used accurately. One of the main duties of a university is to produce graduates ready for their future work. To get familiar with working with a CAD tool, the user need to spend tens or hundreds of hours with using the tool. This is not in correspondence with the time dotation for each subject. It is desirable, that the students will spend extra time working with the software tools. Sometimes, the software developers allow using their software tools for education purposes with some limitations. But unfortunately, this is not a rule and there exists a lot of tools, which cannot be used legally, nor for education purposes. The high price of these tools excludes the option of buying the tools by students for their own needs.

Some software tools are available at the school in limited number of licenses, because of the high price of it. This fact negatively affects the time amount, when the student can work with the software during the lesson.

On the one side, there is a problem with the time space, when the software tools can be used by students, on the other side, the software tools are not used effectively, because in weekly point of view, a huge amount of time, the computers are turned off for several reasons, e.g. the classroom is used for lecture, there is no free stuff, who can take charge of the students in the laboratory, etc.

The solution is to set up remote desktop access to the computers in the classroom, on which the software tools are installed. This will increase the ability to use the computers and their software in 24/7 regime. The only additional cost is the electrical energy, which will be consumed by the computers running the whole day. This can be reduced by waking on the computers only in case of need.

V. Reservation system for remote access

It is not enough to only set up remote access to the desktop of the computers. Very easily can occur a situation, that every computer will be busy, when the user would like to use it. It is necessarily to make a plan of usage, which will guarantee the accessibility of the computer in selected time for the user. We also need to manage the resource usage, to increase the chance, that the computer with needed software tool will be accessible. We also need to guarantee, that the computer will be accessible during the lessons in the classroom and we will not need to abort the remote work of the student to use the computer in the class.

A. Software installation plan

To increase the effectivity of the software usage, we need to carefully create an installation plan, which software tool will be installed on which computer. The ideal situation from the availability's point of view is the case, when the laboratory equipment and every software tool have its own dedicated computer. We need to manage the effective resource allocation on limited number of computer.

On the one side, if we have enough number of licenses to install the tool on every computer, which we would like to make accessible remotely, there is nothing to do. On the other

side, we need to select carefully, on which computers will be the tools installed, which we have available in limited number of licenses. We need to avoid the situation, when a computer will be blocked by user, who is using a tool, which is available on other computer too, and will blocking the usage of tool with limited availability.

For this reason, we need to carefully decide, where to install each license. We need to consider about, who will use the tools, how many students, etc. We need to acquire, that at least one copy of each tool should be available.

B. Software and equipment reservation instead of computer reservation

The reservation system will guarantee exclusive access to the selected computer on which the software tool is installed or which is attached to the laboratory equipment which the user would like to use. For effective usage of the resources, it is important to allow planning the time of usage of each computer.

The reservation system allows for the users to select, when they will have reservation for selected software tool or laboratory equipment. The reservation system, depending on other reservations, will select the most relevant computer and reserve it for the user. The order, in which the computers are allocated, depends on the software, which is installed on them. Firstly, the computers with the software, which is widely available, are allocated.

The computer with the software, which we have available only with one license is priory reserved for this software and is allocated for other software only in case, when no other computer is free. This algorithm allows using the computers much more effectively, than without this logic.

A successful reservation can be created in a few steps. At first, the user must select a free timeslot, when he wants to use the selected tool/equipment. The administrator can accept or reject the reservation. During the active reservation, only the holder of the reservation can access the selected computer. After the reservation expires, the user is automatically logged off. A few minutes before the reservation expires, the user is warned to save his work. After the reservation expires, the computer is ready to use for the next reservation.

The reservation system allows the administrator to set up some limits, e.g. number of reservations for a person for selected time interval; the minimal and maximal length of each reservation; the time interval, when the computer is available for reservations; automatic acceptance of reservations; etc. This ability allows customizing the behavior of the reservation system and adapting it for different needs.

VI. Conclusion

The described reservation system for sharing laboratory equipment and software resources have a big potential to increase the quality of education at the university, because it increases the time that the students can spend on working with equipment and software tools. It has also the potential to upgrade the cooperation between the departments by lending software tools for colleagues from other departments. It should save money for special software tools, which are not needed to use every day, but could be needed to solve some tasks, for which the department has no time and money to purchase it.

Acknowledgment

This paper was developed with support of the project "Centrum excelentnosti integrovaného výskumu a využitia progresívnych materiálov a technológií v oblasti automobilovej elektroniky" (Centre of Excellence of the Integrated Research & Exploitation the Advanced Materials and Technologies in the Automotive Electronics"), ITMS 26220120055, that is co-financed from Structural Funds EU ERDF within Operational Program Research and Development OPVaV-2009/2.1/03-SORO and preferred axis 2 Support of Research and Development.

References

[1] XPort, Embeded Ethernet Device Server, http://www.lantronix.com/device-networking/embedded-device-servers/xport.html

[2] Com Port Redirector, http://www.lantronix.com/device-networking/utilities-tools/com-port-redirector.html

Aging of Polypropylene Films Studied by Solid State ^{1}H NMR

[1]*Peter Duranka (3st year),* [1]*Magdaléna Uhrínová (3nd year)*
Supervisor: [3]*Dušan Olčák*

[1]Dept. of Physics, FEI TU of Košice, Slovak Republic

peter.duranka@tuke.sk, magdalena.uhrinova@tuke.sk

***Abstract*—Aging of biaxially oriented polypropylene films used in packaging industry was studied by means of solid-state ^{1}H MAS NMR. The spectra were measured for the samples stored at room temperature for 15, 150 and 620 days. The changes in mobility of ^{1}H nuclei in different morphological regions with aging time were deduced from the shape alterations of ^{1}H NMR spectra measured at 98ºC. The aging process was found to restrict the chain mobility within non-crystalline regions of the polypropylene films. Densification of the polymer owing to the rearrangement of the chains in non-crystalline regions followed by the free volume decrease was proposed as the reason of the restriction of chain motion.**

***Keywords*—isotactic polypropylene, aging, ^{1}H solid-state NMR.**

I. Introduction

Most commercial polypropylenes are isotactic polypropylenes (iPP) whose chains may be ordered into regions with different arrangement and mobility. Microstructure, physical properties and also commercial applications of iPP depend on the conditions of its preparation and on its thermal history. The iPPs are unique materials, which are frequently used in packaging, textile, electrical, automotive, and many other industries [1-3].

Physical aging of semi-crystalline polymers like iPP is considered as an evolution towards thermodynamic equilibrium manifested by changes in morphology and physical properties. The kinetics of physical aging is given by the initial morphology of polymer and aging temperature. The observed changes of macroscopic properties induced by aging are a consequence of microstructural changes within the semi-rigid and amorphous fractions comparable to those observed after secondary crystallization [4, 5].

This research was stimulated by the requirements from practice. The biaxially oriented iPP films were studied due to the changes of their physical properties after a certain period. These changes can be considered as a consequence of physical aging.

The aim of our study is to evaluate the influence of aging on the molecular structure and dynamics of iPP films stored at room temperature. Structure of polymers determines their properties and molecular motion and NMR spectroscopy is considered to be a unique method for the study of structure and molecular dynamics in solids [4-6]. The magic angle spinning (MAS) ^{1}H NMR is one of the solid state NMR techniques which was picked out for this study.

II. Experimental

A. Studied samples

The ^{1}H MAS NMR experiments were performed on as-supplied biaxially oriented iPP film produced by the Chemosvit a.s., Slovakia. ^{13}C NMR spectra measured on this film, which were reported in previous paper [7] indicate the presence of iPP crystallizing in monoclinic α-form. The NMR measurements were performed on the samples stored in the NMR laboratory at the stable temperature of 23 °C up to 620 days. The aging time t_a in days is indicated in the labeling of the samples as iPP-t_a. Film strips of about 12 mm width were cut and closely rolled before inserting into the rotor.

B. Experimental conditions

The ^{1}H MAS NMR measurements were performed using a Varian solid state NMR spectrometer working at the resonant frequency of 400 MHz at ambient probe temperature, which was of 30 °C and at the temperature of 98 °C. The experiments were carried out with a probe-head using the 4 mm rotor under the magic angle spinning (MAS) conditions at the spinning rate of 10 kHz. The chemical shifts were referenced to the TMS using adamantane as an external standard. The duration of the 90° radio-frequency pulse was 2.7 μs and the recycle delay was 5 s.

III. Results and discussion

The wide-line solid-state ^{1}H NMR spectrum of iPP is determined by the presence of the chemical shift anisotropy and homonuclear dipolar interactions. The spectrum consists of broad, intermediate and narrow lines related to the chains with different mobility in different morphological regions of the partially crystalline polymer [8]. Using the MAS technique, the narrower lines in the spectrum with symmetrically positioned spinning sidebands, arising from the fact that the span of chemical shift anisotropy is larger than the MAS rate, is observed. However, the lines are still broad due the fact that the complete averaging of homonuclear dipolar interactions requires spinning rates higher than the dipolar

linewidth of the static sample and this means the spinning rate of about 40 kHz for iPP at the room temperature [9, 10].

The strong narrowing of the line assigned to the amorphous regions of pelletized iPP was observed with increasing temperature as the result of chain mobility increase in the amorphous regions of iPP structure. A splitting of the central peak into three peaks was observed in the spectra of metallocene and Ziegler-Natta iPPs with the rise of the temperature in the temperature range above 60 °C and 75 °C, respectively [10, 11]. The splitting was alike as in liquid ^{1}H NMR of iPP dissolved in 1, 2-dichlorbenzene [3].

The normalized ^{1}H MAS NMR spectra measured for the iPP-15, iPP-150 and iPP-620 samples at 30 °C are plotted in Fig. 1. These spectra are poorly resolved due to the proton dipolar interactions which are not fully averaged at the used MAS rate, which was 10 kHz. The broad signal of iPP with chemical shift of about 1 ppm is observed in the spectra. Moreover, a very narrow signal (labeled with asteriks in figures) of some substance used probably in manifacturing process of iPP films can be observed in the spectra at 0.1 ppm. Symmetrically positioned spinning sidebands can be clearly recognized in the spectra.

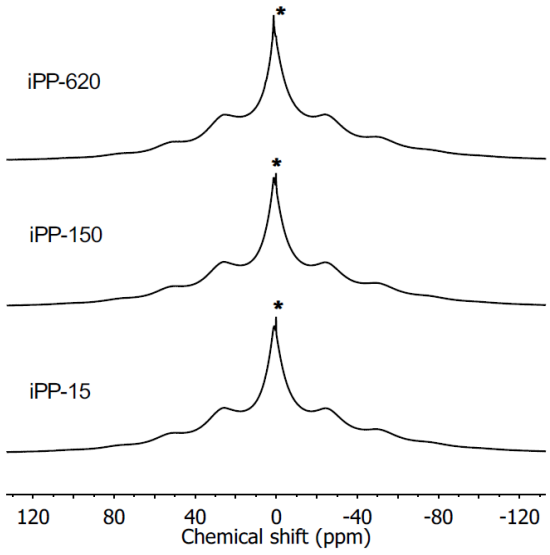


Fig. 1 The normalized ^{1}H MAS NMR spectra measured at 30 °C

Since better resolution of the ^{1}H MAS NMR spectra of pelletized iPP was achieved at elevated temperature [10, 11] we carried out the NMR measurements on iPP films also at the temperature of 98 °C. The spectrum measured at 98 °C can be compared with that measured at 30 °C in Fig. 2 (the spectra are not normalized) and it can be seen that the spectra differ in the shape and width. The strong narrowing of the line assigned to the protons in amorphous regions of the polymer observed in the spectrum recorded at 98 °C (Fig.2 bottom), results from the chain mobility increase in the amorphous regions. The central peaks of normalized ^{1}H MAS NMR spectra measured at the temperature of 98 °C for the iPP films aged for 15, 150 and 620 days are shown in detail in Fig. 3. A splitting of the central peak into three peaks was observed. Using the peak assignments in the liquid ^{1}H NMR spectrum of dissolved iPP [3], the peak at 1.6 ppm can be assigned to the methine protons, the peak at around 1.3 ppm can be assigned to the protons in the methylene group, which are in the anti-position to the adjacent methyl groups and the resonance at 0.9 ppm results from the overlapping of the methyl resonance with the resonance of methylene protons in the syn-position to the adjacent methyl groups. The signals related to the protons in particular chemical groups of iPP broaden with aging time. The line at 1.3 ppm associated with anti-positioned CH_2-protons can be observed only as a left hand shoulder of the line positioned at 0.9 ppm in the spectrum of the sample with the longest aging time. Broadening of the lines may be a consequence of the decreasing mobility in the amorphous regions of iPP with aging. The peak at a chemical shift of 0.1 ppm mentioned above is still present at the spectra but its resolution decreases with aging time.

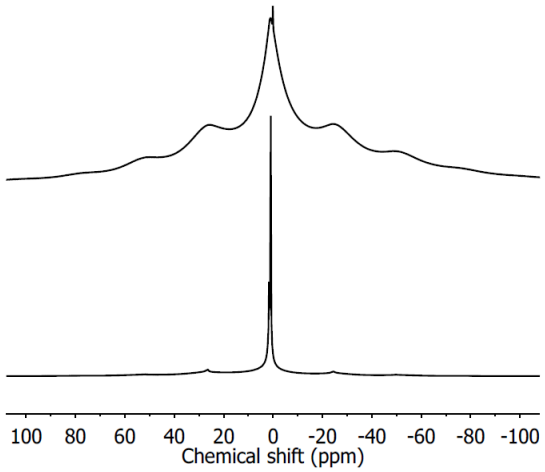


Fig. 2 ^{1}H MAS NMR spectra measured at the sample iPP-15 at 30 °C (up) and iPP-20 at 98 °C (down)

The spectra depicted in Fig. 3 were deconvoluted using three Lorentzian peaks positioned at 1.6, 1.3 and 0.9 ppm related to the protons of individual chemical groups of iPP in mobile amorphous chains and one broad Gaussian peak centered at about 1.3 ppm associated with the protons in the groups of the noncrystalline chains with restricted mobility (less mobile protons) [8]. The widths of the peaks are summarized in Tab. 1.

Deconvolution of the ^{1}H MAS NMR spectra measured at the temperature of 98 °C shows increasing lines broadening with aging time. From this can be deduced that aging process restricts the chains motion within non-crystalline regions probably due to the rearrangement of the chains in these regions resulting in the densification of the polymer structure accompanied by the free volume decrease [4, 5].

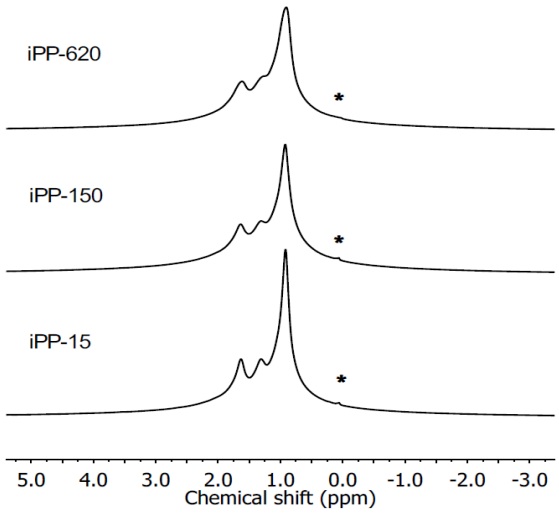


Fig. 3 The normalized ^{1}H MAS NMR spectra measured for the studied samples at 98 °C

Table 1 Widths of the peaks related to the mobile and rigid regions of the samples obtained from the deconvolutions of ^{1}H MAS NMR spectra measured at 98 °C

regions		mobile		rigid
δ (ppm)	1.6	1.3	0.9	1.3
	Width (Hz)	Width (Hz)	Width (Hz)	Width (Hz)
iPP-15	75	90	75	1015
iPP-150	85	105	95	1175
iPP-620	95	115	105	1300

IV. Conclusions

Aging of biaxially oriented iPP films used in packaging industry was studied by means of ^{1}H MAS NMR technique. While the spectra measured at 30 °C do not change significantly with aging time, the spectra measured at 98 °C reveal restriction of chain motion in non-crystalline regions of iPP.

References

[1] M. RATZSCH, "Special PP's for a Developing and Future Market,"„*Journal of Macromolecular Science*, *Part A: Pure Applied Chemistry*, vol. 36, no. 11, 1999, pp. 1587-1611.

[2] D. W. VAN KREVELEN, "*Properties of polymers*," 3rd ed., Amsterdam, Elsevier, 1997, pp. 36-38.

[3] V. BUSICO, R. CIPULLO, " Microstructure of Polypropylene," *Progress in Polymer Science*, vol. 26, 2001, pp. 443-533.

[4] E. R. SADIKU, N. PEACOCK, "Physical Aging of Polypropylene Filaments by Density Measurements," *Kautschuk und Gummi, Kunststoffe*, vol. 43, no. 1, 1990, pp. 26-28.

[5] C. HEDESIU, D. E. DEMCO, R. KLEPPINGER, G.V. POEL, K. REMERIE, V.M. LITVINOV, B. BLÜMICH, R. STEENBAKKERS, "Aging Effects on the Phase Composition and Chain Mobility of Isotactic Poly(propylene), "*Macromolecular materials and Engineering*, vol. 293, no. 10, 2008, pp. 847-857.

[6] F. A. BOVEY, P.A. MIRAU, "*NMR of Polymers*," San Diego, Academic Press, USA, 1996, pp. 243-352.

[7] O. FRIČOVÁ, D. OLČÁK, M.KOVAĽAKOVÁ, V. HRONSKÝ, "Aging of polypropylene foils studied by solid state NMR," *International conference on applied electrical engineering and informatics*, Italy, pp. 70-74, Sept. 2011.

[8] C. HEDESIU, D.E. DEMCO, R. KLEPPINGER, G.V. POEL, B. BLÜMICH. K. REMERIE, V.M. LITVINOV, "Effect of Temperature and Annealing on the Phase Composition, Molecular Mobility, and Thickness of Domains in Isotactic Polypropylene Studied by Proton Solid-State NMR," SAXS and DSC, *Macromolecules*, vol. 40, 2007, pp. 3977-3989.

[9] M. J. DUER, "*Introduction to Solid-State NMR Spectroscopy*," Oxford, Blackwell Publishing, (2004), pp. 61-77.

[10] M. UHRÍNOVÁ, O. FRIČOVÁ, D. OLČÁK, V. HRONSKÝ, "^{1}H MAS NMR study of molecular motion in isotactic polypropylene," *16th international conference on Applied Physics of Condensed Matter*, MaláLučivná, Slovakia, pp. 179-182, June 2010.

[11] O. FRIČOVÁ, M.UHRÍNOVÁ, V. HRONSKÝ, M. KOVAĽAKOVÁ, D. OLČÁK, I. CHODÁK, J. SPĚVÁČEK,"High-resolution solid-state NMR study of isotactic polypropylenes," *Express Polymer Letters*, 2012, Vol. 6, 204-212.

An Approach for Determination the Mutual Influence of the Supply Mains and Non-Linear Load

[1]*Dmytro MAMCHUR (5th year),* [2]*Matúš OCILKA (2nd year)*
Supervisor: [3]*Andrii KALINOV*

[1,3]Institute of Electromechanics, Energy Saving and Control Systems, Kremenchuk Mykhailo Ostrohradskyi National University, Ukraine
[2]Dept. of Theoretical Electrical Engineering and Electrical Measurement, FEI TU of Košice, Slovak Republic

[1]scenter@kdu.edu.ua, [2]matus.ocilka@tuke.sk

***Abstract* – The approach to determination of mutual influence of the consumer and supply mains is considered. The analysis of the possibility of usage the instantaneous power signal spectrum analysis for separation the voltage harmonics generated by power system and harmonics generated by non-linear consumer in a connection point is shown.**

***Keywords* – Diagnostics, instantaneous power, power system.**

I. Introduction

Nowadays the most progressive methods for determination the IM real technical conditions based on the analysis of currents, voltages and powers instantaneous values [1]. These signals should be measured on the consumer connectors, in our case – ob IM connectors. It should also taken into account that both supply mains low quality and consumer non-linearity influence on harmonic composition of the mentioned signals. IM diagnostics in field conditions supposed analyzing motor signals fed by low-quality supply. The influence of the mains low-quality decreases accuracy and distorted results of the diagnostic. Aim of research is the extraction the contribution of supply mains low-quality into harmonic composition of current and voltage signal in tasks of IM analysis.

II. Research results

For separation influence of supply mains low-quality on electrical signal in connection point, we analyzed equivalent circuit "Supply mains – non-liner consumer" (fig. 1).

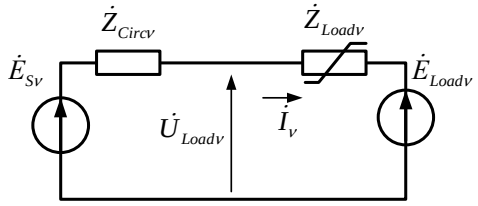


Figure 1 – Equivalent circuit "Supply mains – non-linear consumer"

Our task could be solved if we represent supply mains as a linear object with lumped parameters of active-inductive character, which is correct for the mains under 1000 V. The energy stream direction of the certain harmonics could be determined by analyzing the constant component of the instantaneous power signal [2]. Let compose voltage balance equations for the given equivalent circuit:

- $\dot{U}_{Loadv} = \dot{E}_{Sv} - \dot{I}_v \dot{Z}_{Circv}$ – for the case of power supply harmonics domination;
- $\dot{U}_{Loadv} = \dot{E}_{Sv} + \dot{I}_v \dot{Z}_{Circv}$ – for the case of non-liner consumer harmonics domination;

where v is the harmonic number, $\dot{U}_{Loadv}$ is the voltage harmonic on the consumer connector, $\dot{E}_{Sv}$ and $\dot{E}_{Loadv}$ are emf harmonics of power supply and non-linear load, $\dot{I}_v$ is the current harmonic, $\dot{Z}_{Sv}$ and $\dot{Z}_{Loadv}$ are full complex resistances of equivalent circuit and non-linear consumer.

In case of unknown values of supply emf $\dot{E}_{Sv}$ and mains parameters $\dot{Z}_{Sv} = R_S + jv\omega L_S$ the solution of this task is impossible. However, solution could be find using instantaneous power balance equations [2]:

$$\begin{cases} P_{0Load} = P_{0S} \pm P_{0R_C}; \\ P_{kaLoad} = P_{kaS} \pm \left(P_{kaL_C} + P_{kaR_C} \right); \\ P_{kbLoad} = P_{kbS} \pm \left(P_{kbL_C} + P_{kbR_C} \right). \end{cases} \quad (1)$$

– for power supply:

$$Pa_{mS} = 0.5\left(\begin{array}{l}\sum\limits_{\substack{k=0\\m-k\geq 0}}^{N-1} Ia_k Ea_{m-k} + \sum\limits_{\substack{k=0\\m-k<0}}^{N-1} Ia_k Ea_{k-m} - \\ - \sum\limits_{\substack{k=0\\m-k\geq 0}}^{N-1} Ib_k Eb_{m-k} + \sum\limits_{\substack{k=0\\m-k<0}}^{N-1} Ib_k Eb_{k-m}\end{array}\right);$$

$$Pb_{mS} = 0.5\left(\begin{array}{l}\sum\limits_{\substack{k=0\\m-k\geq 0}}^{N-1} Ib_k Ea_{m-k} + \sum\limits_{\substack{k=0\\m-k<0}}^{N-1} Ib_k Ea_{k-m} + \\ + \sum\limits_{\substack{k=0\\m-k\geq 0}}^{N-1} Ia_k Eb_{m-k} - \sum\limits_{\substack{k=0\\m-k<0}}^{N-1} Ia_k Eb_{k-m}\end{array}\right);$$

– for power on consumer connectors:

$$Pa_{mLoad} = 0.5\left(\begin{array}{l}\sum\limits_{\substack{k=0\\m-k\geq 0}}^{N-1} Ia_k Ua_{m-k} + \sum\limits_{\substack{k=0\\m-k<0}}^{N-1} Ia_k Ua_{k-m} - \\ - \sum\limits_{\substack{k=0\\m-k\geq 0}}^{N-1} Ib_k Ub_{m-k} + \sum\limits_{\substack{k=0\\m-k<0}}^{N-1} Ib_k Ub_{k-m}\end{array}\right);$$

$$Pb_{mLoad} = 0.5\left(\begin{array}{l}\sum\limits_{\substack{k=0\\m-k\geq 0}}^{N-1} Ib_k Ua_{m-k} + \sum\limits_{\substack{k=0\\m-k<0}}^{N-1} Ib_k Ua_{k-m} + \\ + \sum\limits_{\substack{k=0\\m-k\geq 0}}^{N-1} Ia_k Ub_{m-k} - \sum\limits_{\substack{k=0\\m-k<0}}^{N-1} Ia_k Ub_{k-m}\end{array}\right);$$

– for supply active resistance:

$$Pa_{mR} = 0.5R_C\left(\begin{array}{l}\sum\limits_{\substack{k=0\\m-k\geq 0}}^{N-1} Ia_k Ia_{m-k} + \sum\limits_{\substack{k=0\\m-k<0}}^{N-1} Ia_k Ia_{k-m} - \\ - \sum\limits_{\substack{k=0\\m-k\geq 0}}^{N-1} Ib_k Ib_{m-k} + \sum\limits_{\substack{k=0\\m-k<0}}^{N-1} Ib_k Ib_{k-m}\end{array}\right);$$

$$Pb_{mR} = 0.5R_C\left(\begin{array}{l}\sum\limits_{\substack{k=0\\m-k\geq 0}}^{N-1} Ib_k Ia_{m-k} + \sum\limits_{\substack{k=0\\m-k<0}}^{N-1} Ib_k Ia_{k-m} + \\ + \sum\limits_{\substack{k=0\\m-k\geq 0}}^{N-1} Ia_k Ib_{m-k} - \sum\limits_{\substack{k=0\\m-k<0}}^{N-1} Ia_k Ib_{k-m}\end{array}\right);$$

– for supply inductance:

$$Pa_{mL_C} = \frac{L_C m\omega}{4}\left(\begin{array}{l}-\sum\limits_{\substack{k=0\\m-k\geq 0}}^{N-1} Ib_k Ia_{m-k} - \sum\limits_{\substack{k=0\\m-k<0}}^{N-1} Ib_k Ia_{k-m} - \\ - \sum\limits_{\substack{k=0\\m-k\geq 0}}^{N-1} Ia_k Ib_{m-k} + \sum\limits_{\substack{k=0\\m-k<0}}^{N-1} Ia_k Ib_{k-m}\end{array}\right);$$

$$Pb_{mL_C} = \frac{L_C m\omega}{4}\left(\begin{array}{l}\sum\limits_{\substack{k=0\\m-k\geq 0}}^{N-1} Ia_k Ia_{m-k} + \sum\limits_{\substack{k=0\\m-k<0}}^{N-1} Ia_k Ia_{k-m} - \\ - \sum\limits_{\substack{k=0\\m-k\geq 0}}^{N-1} Ib_k Ib_{m-k} + \sum\limits_{\substack{k=0\\m-k<0}}^{N-1} Ib_k Ib_{k-m}\end{array}\right);$$

where Ia_k, Ia_{m-k}, Ib_k, Ib_{m-k} are cosine and sine components of current harmonics of the orders k and $m-k$;

Ea_k, Ea_{m-k}, Eb_k, Eb_{m-k}, Ua_k, Ua_{m-k}, Ub_k, Ub_{m-k} are cosine and sine components of e.m.f. and load voltage harmonics, respectively.

Solving this system of equations some uncertainty appears. The system has infinite set of solutions. To find required solution, we have to know beforehand at least one searching value. As such value, we can preliminary find supply parameters R_C and L_C. For this, let investigate circuit "power supply – active load" (fig. 2).

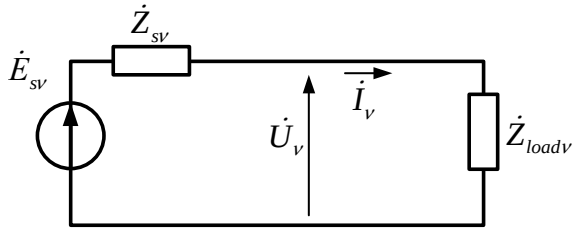


Figure 2 – Equivalent circuit "power supply – active load"

For circuits with non-sinusoidal power supply of mains frequency it considered that the resistances are not depending on frequency. At the same time, the inductance growths directly proportional to the frequency [3].

Thus, total active resistance and inductance of the whole circuit could be determined by the following expressions:

$$R_{circ} = Re\left(\frac{\dot{E}s_1}{\dot{I}_1}\right);$$

$$L_{circ\nu} = \frac{1}{2\pi f\nu} Im\left(\frac{\dot{E}s_1}{\dot{I}_1}\right),$$

where $\dot{E}_{s_1}$, $\dot{I}_1$ are e.m.f. of first harmonics of the supply and current in complex form, respectively; f is the supply mains frequency; ν is the harmonic number.

In real conditions current and voltage sensors could be set up in consumer connection point. Measured current will be equal for all circuit. Thus, we can find total active resistance and inductance of the consumer:

$$R_{cons} = Re\left(\frac{\dot{U}_1}{\dot{I}_1}\right);$$

$$L_{cons} = \frac{1}{2\pi f\nu} Im\left(\frac{\dot{U}_1}{\dot{I}_1}\right),$$

where $\dot{U}_1$ is the first voltage harmonic on consumer side in a complex form.

According to presented equal circuit, mains active resistance and inductance we can get in the following way:

$$R_C = R_{II} - R_{II} = Re\left(\frac{\dot{E}u_1}{\dot{I}_1}\right) - Re\left(\frac{\dot{U}_1}{\dot{I}_1}\right);$$

$$L_C = L_{II} - L_{II} = \frac{1}{2\pi f \nu}\left(Im\left(\frac{\dot{E}u_1}{\dot{I}_1}\right) - Im\left(\frac{\dot{U}_1}{\dot{I}_1}\right)\right).$$

In real conditions while determining described values unknown parameters, except for mains resistances also are mains e.m.f. harmonics values. Thus, to determine all unknown values we need to get additional equations. This could be done by changing consumer resistances. For determination supply mains parameters it proposed connect to supply mains turn by turn resistors with determined resistances and measure instantaneous currents and voltages on connectors. In this case we'll get the following system of equations:

$$\begin{cases} R_{mains} = Re\left(\frac{\dot{E}s_1}{\dot{I}1_1}\right) - Re\left(\frac{\dot{U}1_1}{\dot{I}1_1}\right); \\ R_{mains} = Re\left(\frac{\dot{E}s_1}{\dot{I}2_1}\right) - Re\left(\frac{\dot{U}2_1}{\dot{I}2_1}\right); \\ \vdots \\ R_{mains} = Re\left(\frac{\dot{E}s_1}{\dot{I}n_1}\right) - Re\left(\frac{\dot{U}n_1}{\dot{I}n_1}\right), \end{cases}$$

$$\begin{cases} L_{mains} = \frac{1}{2\pi f \nu}\left(Im\left(\frac{\dot{E}s_1}{\dot{I}1_1}\right) - Im\left(\frac{\dot{U}1_1}{\dot{I}1_1}\right)\right); \\ L_{mains} = \frac{1}{2\pi f \nu}\left(Im\left(\frac{\dot{E}s_1}{\dot{I}2_1}\right) - Im\left(\frac{\dot{U}2_1}{\dot{I}2_1}\right)\right); \\ \vdots \\ L_{mains} = \frac{1}{2\pi f \nu}\left(Im\left(\frac{\dot{E}s_1}{\dot{I}n_1}\right) - Im\left(\frac{\dot{U}n_1}{\dot{I}n_1}\right)\right), \end{cases}$$

where $\dot{I}1_1, \dot{I}2_1, \ldots, \dot{I}n_1$; $\dot{U}1_1, \dot{U}2_1, \ldots, \dot{U}n_1$ are first harmonics of measured in circuit currents and voltages for different values of tested resistances. Solving this system, we can determine R_C and L_C.

Placing determined values into system (1) we can determine values of all the rest supply e.m.f. harmonics.

For checking theoretical researches, the experiment was done. In Matlab environment the simulation of circuit, which contains non-sinusoidal power supply (the combinations of 1st, 3rd, 5th and 7th harmonics were defined) and consumer, which is characterized by active resistance and inductance and contain harmonics source which generates harmonics with different amplitudes (higher or lower then supply mains harmonics), was done. Part of harmonics, generated by consumer, has higher amplitudes then power supply harmonics (fig. 3). According to scheme, either circuit with changeable resistances or circuit with consumer described by active resistance, inductance and harmonics source, could be connected. For simulating signals parameters close to real conditions the source of noise was included in circuit with noise amplitudes lower 1% of tested signals amplitudes.

Initially, scheme connects turn b turn with different testing resistances in order to define supply mains parameters. So as for the following given supply mains parameters as $R_{mains} = 1{,}35\ Ohm$ and $L_{mains} = 0{,}012\,Ohm$, and e.m.f. supply first harmonic of $\dot{E}_i = 0{,}24506 - 155{,}53303i$ solution of composed system of equations for four tested resistances we can get following values: $R_{mains} = 1{,}34995\ Ohm$, $L_{mains} = 0{,}0120137\ Ohm$, $\dot{E}_i = 0{,}2450313 - 155{,}5330047i$, which confirm accuracy of proposed method.

After this, scheme connects to loading circuit. Basing on measures, a system of equations according to (1) could be composed. Solutions of this system for case of existence the 1st, 3rd, 5th and 7th harmonics of power supply and 1st, 3rd, 5th, 7th and 9th harmonics generated by consumer are presented in table 1.

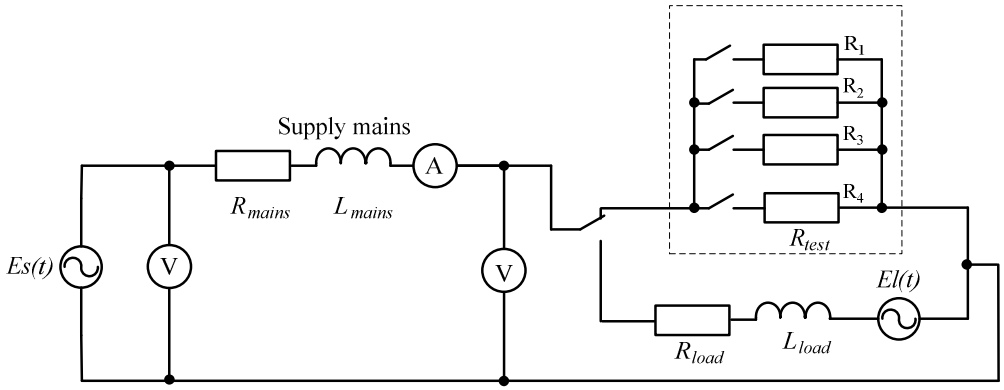


Figure 3 – Scheme of analyzed equivalent circuit

Table 1 – Evaluation of experimental results

Given values of emf harmonics	Solutions of the system of equations	Accuracy, %
$\dot{E}_{i3} = 0{,}13985 - 31{,}07532i$	$\dot{E}_{i3} = 0{,}140626 - 31{,}07557i$	0,6
$\dot{E}_{i5} = 0{,}11216 - 15{,}50546i$	$\dot{E}_{i5} = 0{,}11094 - 15{,}51176i$	1,1
$\dot{E}_{i7} = 0{,}08702 - 7{,}72636i$	$\dot{E}_{i7} = 0{,}09446 - 7{,}725i$	8,5

Aberration of solution could be explained by noise influences on current and voltage signal and also by computational aberrations. It also should be mentioned that composing system of equations we didn't take into account balance equations for 9th harmonic because its insignificantly amplitude value comparable to noise value.

III. Conclusion

A method for determination the supply mains parameters based on currents and voltages analysis on the connectors of additional resistances was developed. An algorithm for determination of power supply harmonics numeric values based on determined supply mains parameters also was developed. Calculated values could be further used for reduction the polyharmonic signal of consumer on connectors to sinusoidal signals. This will allow analyzing appearance and development of distortions injected by consumer into electrical signals and could be used for tasks of monitoring and diagnostics of electromechanical equipment.

IV. References

[1] Rodkin D.I., "A new system of the quality rates of electrical energy usage,"Scientific herald of The National Mining University, 2004, pp. 20 – 26. (in Russian)

[2] Kalinov A.P., Mamchur D.G., Angulo J.V., "Separation of the supply mains low-quality influence on AC electrical motors in diagnostics , The Herald of National Technical University "KhPI", "Problems of automated electric drive. Theory and practice" – Kharkiv, NTU "KhPI", 2008, pp. 559 – 563. (in Ukrainian)

[3] Bessonov L.A., "Theoretical basics of the electrotechnics. Electrical circuits: textbook," Moscow: Gardariki, 2001 , 638 p. (in Russian)

Analysis of the dielectric absorption by maximum likelihood estimation

[1]*Marek GODLA(3st year)*, [2]*Martin BRODA (2st year)*
Supervisor: [3]*Linus MICHAELI*

[1,2,3]Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

[1]marek.godla@tuke.sk, [2]martin.broda@tuke.sk

***Abstract*— Nowadays, in the world of digital equipments, it is necessary to use input analog circuits for adjustment of input value of the parameter before using analog-digital convertor. Dielectric absorption has the biggest influence on the quality of the capacitor. This paper gives a proposal for a method of searching of components of Dow`s model of dielectric absorption.**

Searching of components is performed by maximum likelihood estimation and the Least Square Method.

The paper deals with simulated parameters where is simulated recharging of capacitor and components of Dow's model are known.

Keywords – exponential signal, capacitor, absorption, Dow's method, maximum likelihood estimation

I. INTRODUCTION

Dielectric absorption is a tendency of a capacitor to recharge itself after being discharged. While not widely appreciated dielectric absorption is also the dominate loss model over usable range of most capacitors and so it can also be the dominant loss model over the entire usable range of most capacitors. [1]

If a capacitor is charged for a long period of time and then it is shorted, the voltage on the capacitor will slowly tend to recovery on the percentage of input voltage (typically 0.01% to 10%). [1]

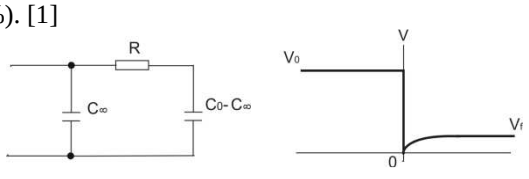


Fig 1. Model of capacitor

The model shows the dielectric absorption aspect of both memory and dissipation.
When a capacitor is charged, dielectric is polarized and creates of dipoles. According to von Hippel [6] the polarization mechanism can be divided into four different types.

1. Electronic polarization – upon the application of an external electric field, the electron clouds of the nuclei are displaced slightly
2. Atomic polarization – there often are asymmetries in the molecules formed by atoms of different types.
3. Orientation polarization – even without an electric field molecular dipoles are created
4. Space charge polarization - a dielectric always contains a small number of more or less mobile charge carries.

II. DOW'S METHOD

A method presented by Charles H. Dow is used for the time of the domain modeling of dielectric absorption. According to his method a capacitor is charged for a long time and suddenly shorted. This voltage can be approximated by a summation of an exponential signal,

$$\frac{I}{V_0} = C_\infty \sum_{k=1}^{n} a_k e^{\frac{t}{\tau_k}} \qquad (1)$$

where C_∞ represents a capacity of a real capacity of the capacitor, τ_l and a_k are constants of the exponentials. The exponentials are real and can be modeled as RC pairs. We are summing current I the nehodí sa v článku priebehový tvar shorted circuit and therefore RC pairs will be added in parallel to C_∞.

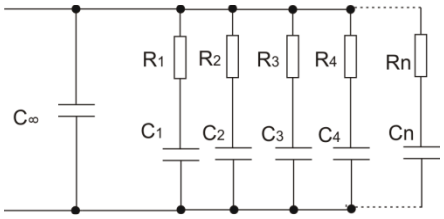


Figure 2. Absorbive model

The figure 2 shows the distortion of dielectric by RC pairs which are undesirable in the real circuits. Values of these components are unknown for the designers of circuits.
This paper suggests a method for measuring these components.

III. NEW METHOD FOR SEARCHING OF EXPONENTIALS

The Dow's model gives a mathematical description of the distortion by a summation of exponential signals. Practically

it is impossible to obtain every exponential signal separately, therefore several methods for measuring of every exponential in the final output exponential signal exist. Identification of the exponentials based on FFT spectrum analysis is not suitable in the case when components aren´t harmonic or orthogonal. The spectrum is continuous in this case and the parameters of components cannot be identified.

The exponential components are clearly not orthogonal. This is why the estimation uses the traditional methods based on the least squares (LS) or algorithms are sensitive to the noise distortion [2] and to the number of recorded samples from the ADC output. An alternative way of analysis is the identification of the superimposed components by using Prony's estimation method, in which sampling is equidistant [4].

The Maximum likelihood (ML) estimation is proposed and studied further under the consideration of analog to digital conversion with low nonlinearity. In contrary to the identification of the distorted harmonic signal for the dynamic ADC test [5], the non orthogonal exponential components are identified by using ML method under the consideration of the ideal AD conversion. The ML method has increased computational complexity, but in contrast to the LMS method it has a low sensitivity to clipping of the signal and to nonlinearity of the ADC. The computation time will be increased because of the multidimensional optimization. The selection of the initial values in the optimization algorithm can speed up the whole procedure. The initial conditions and influence of superimposed noise will be studied by the simulation developed in the LabVIEW environment. The digital data from the ADC output will be acquired using the ADC model with the chosen parameters and an adopted INL function.

IV. MATHEMATICAL MODEL

In our proposed method we changed equation (1) on the shape that shows the next equation (2),

$$x_{in}(t) = A_1 e^{-B_1 t} + \sum_{i=2}^{n} A_i e^{-B_i t} + C \quad (2)$$

where the parameter A_i represents resistive components and B_i represents capacitance components caused by dielectric absorption. The parameter C describes offset of the whole exponential signal. The components meet the following conditions: $A_1 >> A_i$, $0 < B_1 < B_i$. The distorted multiexponential is measured together with the additional noise generated by the analog components. The noise $n(t)$ is assumed to have Gaussian distribution with zero mean and variance σ^2 .

$$x_s(j) = \sum_{i=1}^{n} A_i . e^{-B_i \cdot jT_S} + C + n(jT_S) \quad (3)$$

We consider analog to digital conversion with an ideal ADC. The quantization levels T(k) are equidistant T(k) = Q(k+0.5), where Q is the ideal code bin width. The recorded digital sample from the ADC output $k(j)$ corresponds with the j-th sample of input exponential signal $x_s(j)$.

The best estimation of the excitation signal is represented by the situation when the code samples $k(j)$ match best the values of the sampled input signal $x(j)=x_s(j)/Q$, normalized by the code bin width Q. The probability of any digital sample $k(j)$ matching the normalized input value $x(j)$ is expressed by one of following expressions:

$$P(k(j)=0, x(j)) = \\ = F([T(0)-x(j)],0,\sigma) - F([-\infty - x(j)],0,\sigma) \quad (3)$$

$$P(k(j), x(j)) = \\ = F([T[k]-x(j)],0,\sigma) - F([T(k-1)-x(j)],0,\sigma) \quad (4)$$

$$P(k(j)=(2^N-1), x(j)) = \\ = F([\infty - x(j)],0,\sigma) - F([T(2^N-2)-x(j)],0,\sigma) \quad (5).$$

The optimization task is to find constants A_i, B_i, C from expression (2) when the product of probabilities $P((k(j),x(j))$, for all codes $k(j)$ achieves maximum. The maximum of the joint probability P_{final} represents the maximal likelihood

$$P_{final} = \prod_{j=0}^{I} P((k(j), x(j)) \quad (6).$$

When the optimization works with the equation (X) there will be rounding errors of P_{final} caused by high number.

The most suitable optimisation strategy is to find minimum of the negative logarithm of (6).

$$c\ (p) = -\log P_{final} = -\sum_{j=0}^{I} \log P(k(j), x(j)) \quad (7)$$

V. EXPERIMENTAL RESULTS

Tests were performed with the data acquired using simulated ideal 12-bit ADC. The evaluation of the proposed ML method was the main goal of the performed tests based on the generated parameters and estimated ones. The robustness of the proposed ML method on the number of samples and superimposed noise was studied by the simulation tool developed in the LabVIEW environment. The Figure 1 shows a block structure of the simulation program. The signal from the generator of the multiexponential signal $xs(jT_s)$ is quantized in the ADC model which allows to implement known nonlinearity function INL(k). The number of the exponential components was chosen according to three components.

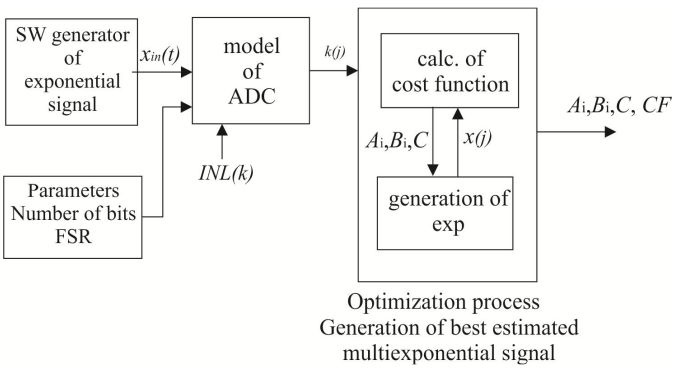


Fig. 3 Diagram of the simulation software to generate exponential signal and fitting by Maximum likelihood estimation

For the optimization was selected differential evolution . The convergence process was accelerated by adaptive changing of the dispersion of Gaussian distribution in (3),(4),(5).
The shape of the simulated multiexponential signal with three exponential components without and with noise is shown in Fig. 4. The effective value of the superimposed noise is equal to 0,8 LSB. The input signal was converted by the 12 bit ADC and for the testing of the optimization efficiency 800 samples was registered.

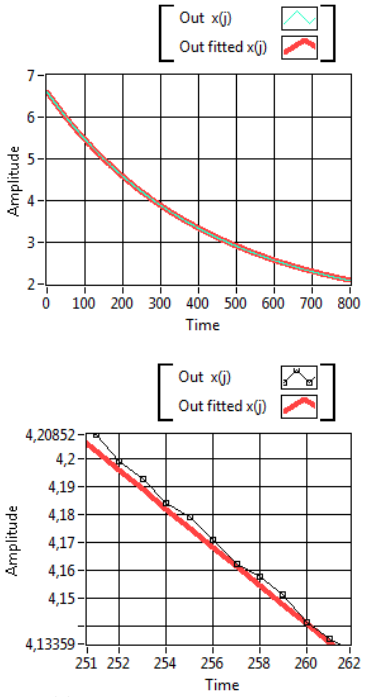


Fig. 4 Results of fitting by Maximum lilkelihood estimation

Simulations showed high sensitivity of the optimization process on the initial conditions. Increasing number of the exponential functions (2) requires good choice of the initial conditions. The optimization procedure converges for any initial values A_i, B_i, C when the number of exponentials is no more than two. The suitable estimation of the initial values is required when the number of exponential is three. The drawback is that the estimation of higher components is strongly influenced by the noise level. Another possibility is utilization of the modified Pony's method [4] for the first estimate of exponential components.

Next the optimization of the cost function (7) was performed. The results of the parameter identification for noisy and noiseless signal are shown in Table 1.

TABLE I
Results of simulations

Constants	Real input	Estimated parameters without noise	Estimated parameters with noise σ=0,8LSB
A1	5	5,0013	5,0074
A2	1	0,9729	1,0646
A3	0,5	0,5238	0,3854
B1	5	4,99817	4,9972
B2	0,50	0,5116	0,4046
B3	0,05	0,0658	0,2035
C	0,1	0,0959	0,1405
Cost funct.	-	699,72	1119,8

Estimated parameters for known input signal for both cases without noise (third column) and with superimposed noise with effective value of 0,8 LSB.

VI. Conclusions

The simulation shows that ML method is more resistant to the influence of the superimposed noise on the identification of the multiexponential signal parameters. The further study will be focused on the influence of number of samples and the efficiency of the optimization procedure. The final results will be approved by the experimentally acquired pulses by the real DAQ system.

References

[1] Kundert, K.:"Modeling Dielectric Absorption in Capacitors", www.designers-guide.org ,2002.
[2] Saliga,J, Michaeli,L., Sakmar,M, Busa,J.: "Processing of bidirectional exponential stimulus in ADC testing" Measurement. Vol. 43, no. 8 (2010), p. 1061-1068. - ISSN 0263-2241
[3] Saliga,J, Michaeli,L., Holcer,R. :"Noise sensitivity of the exponential histogram ADC test", Measurement. Vol. 39, no. 3 (2006), p. 238-244. - ISSN 0263-2241
[4] Carni,D.L. et al.: "Measurement of the exponential signal distortion" IEEE Procedings of I2MTC 2012 : May 13-16, Graz, Austria. - Graz : P. 1773-4. - ISBN 978-1-4577-
[5] Balogh,L., Kollár,I., Sárhegyi, A:"Maximum Likelihood Estimation of ADC Parameters." IEEE Procedings of IMTC 2010. Austin, United States of America, 03/05/2010-06/05/2010. pp. 24-29.
[6] A.R. von Hippel, Dielectric Materials and Applications. New York:Willey, 1954, 18-19
[7] Hyppä K.:"Dielectric Absorption in Memory Capacitor" IEEE Transactions on instrumentation and measurement, February 1972

Analyze of primary controllers and frequency control process in power system

[1]*Pavol HOCKO (3[rd] year),* [2]*Vieroslava SKLENÁROVÁ (3[rd] year)*
[1]*Supervisor1:* [3]*Michal KOLCUN,* [2]*Supervisor2:* [4]*Roman CIMBALA*

[1, 2] Dept. of Electric Power Engineering, FEI TU of Košice, Slovak Republic
[3, 4] Dept. of Electric Power Engineering, FEI TU of Košice, Slovak Republic

[1]pavol.hocko@tuke.sk, [2]vieroslava.sklenarova@tuke.sk, [3]michal.kolcun@tuke.sk, [4]roman.cimbala@tuke.sk

***Abstract* — this paper discuss about frequency stability of power system. The power system is represented as power grid which consists of 30 nodes. Frequency control modeling falls under middle-term dynamics. Modeled power grid was made in MODES simulation software. Different primary controllers were compared to investigate primary frequency regulation process. Also frequency load shedding is simulated to compare different control mode behavior in power system.**

***Keywords*— frequency control, frequency load shedding, middle-term dynamics, MODES, support services, power system stability.**

I. INTRODUCTION

Middle-term dynamics is characterized by time interval from few tens of seconds to minutes. Power supplies drop-outs and load changes are investigated in long-term dynamics. Effects of these faults in the way of power system response and frequency deviation stability are verifying. To ensure power system stability frequency has to be maintained in small limited interval. This is required to secure safety operation of machines on the production side and load on the consumption side. In normal operation of power system the frequency is maintained with Primary Active Power and Frequency Control mechanism. This paper discuss about Primary Frequency Control system and Frequency Load Shedding which is activated in case of lack of power to maintain power balance in power system.

II. MODELED POWER SYSTEM

Power system designed to modeling primary frequency control process will be described in this chapter. Figure 1 shows modeled power system. This power system consists of 30 nodes in three different voltage levels. Whole power system is divided in four control areas. Small areas consisting only one generator represents different power systems connected to our main power grid and are marked with circles. The remaining nodes, lines and transformers which represent main power system are part of control area "Area 1". Eight turbo-generator systems with different parameters supply whole interconnected power system.

Disconnection of generator BLOK3 will be modeled in time of 1 second as is shown in Figure 1. For frequency shedding different scenario will be accepted (BLOK1instead of BLOK3 will be switched off).

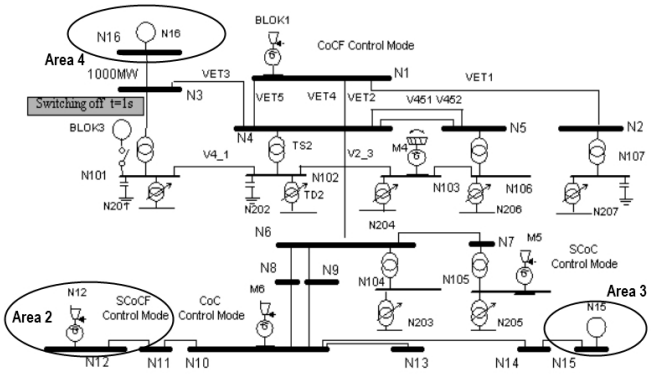


Fig. 1 Connection of modeled power system with marked control areas

III. PRIMARY FREQUENCY AND POWER CONTROL

A. Frequency regulation process

The primary frequency/power control (shorten Primary Regulation) reserves the mechanical power on the generator's turbine. This reserved power is used to maintain the frequency deviation in power system after power balance disruptions. This response has to be achieved in time of seconds and has to be permanent until frequency deviation lasts. Primary Regulation is based on primary control reserve which is reserved on turbine's regulation range. Power reserve has to be fully activated in time of 30 seconds after fault occurs. Primary Regulation has proportional (static) character. Proportional character of Primary Regulation leaves the frequency deviation after a disruption of power balance. Whereas the Primary Regulation is activated on all generators with primary controller, temporary breakdown help is realized in struck control area.

The Primary Regulation may be realized by:

- turbine power controller,
- proportional operation speed controller,
- combination of first two in serial or parallel configuration.

B. *Frequency control mode models*

Each generator in described power system has different primary control mode thus the dynamics of primary control activation will be different. Each control mode will be described below.

Each turbine mechanically connected to generator has own regulator which controls amount of steam (or water) delivered to turbine to control active power which generators delivers to power system. Regulator's control mode is represented by regulation code. Each letter in code describes the parameters and performance of given controller. Parameters corresponding to individual letters are shown in table 1 where first letter represents speed controller, second letter is for turbine regulation regime, third is type of block itself, forth letter indicates the frequency correction.

TABLE I
CONTROL CODES USED IN SIMULATION

1st	2nd	3rd	4th	Function	
S				Speed controller ON	
N/A				Speed controller OFF	
	Po			Classic regulation	Turbine control regime
	Pr			Inlet pressure reg.	
	Co			Coordinated regulation	
	Ma			Manual control	
	Sl			Natural sliding pressure	
		C		Constant pressure	Turbine type
		V		Variable pressure	
		N		Nuclear reactor	
		H		Hydro turbine	
		G		Gas turbine	
			F	Freq. controller ON	Freq
			N/A	Freq. controller OFF	

C. *Primary frequency control analysis in MODES*

Primary control process is simulated in MODES. First step is to analyze primary regulation. MODES can analyze static and dynamic characteristic for individual regulation area and also summary characteristics of all areas. As was mentioned in first chapter the primary regulation process has proportional character with its activation in function of static characteristics (1). Primary regulation analysis calculates the power share from turbines in power system (N_t) as a function of frequency (f) which is changed in steps from minus 1000 mHz to 1000 mHz in step of 50 mHz.

$$N_t = f(\mathit{frequency}) \tag{1}$$

From this analysis it is now possible to compute power number of each power system or resultant power number for whole interconnected power system. Static power number is defined for small frequency deviations generally under 100 mHz and it is given by kCOR coefficient in power regulator parameter settings. According to (2) the static power number for whole power system is 3380 MW/Hz.

$$\lambda = \frac{\Delta P}{\Delta f} = \frac{338\ MW}{0.1\ Hz} = 3380\ MW / Hz \tag{2}$$

Dynamic power number is calculated for a large frequency deviations (above 500 mHz) at this frequency deviation the controller saturation reaches maximum level and turbine power is set by a speed controller thus the kSP coefficient in power regulator is taken to account. According to (3) the dynamic power number for frequency deviation 500 mHz is 2176.8 MW/Hz.

$$\lambda = \frac{\Delta P}{\Delta f} = \frac{899MW}{0.5Hz} = 2176.8MW / Hz \tag{3}$$

It is important to mention that the power number of the power system is not a constant but a variable value. Value of power number depends mainly on the frequency deviation.

D. *Reserved power for primary regulation*

Each power block providing primary power control has defined reserved power for primary control. This reserve represents a percentage share of installed power of generator and its value depends on turbine dynamic properties. This value changes in scale from 0% to 5%. Analyze of primary control power reserve is in table Tab. 2. All values are in MW.

TABLE II
PRIMARY POWER CONTROL ANALYZE

Generator Name	Regulator Code	Rezerve [MW]		Turbine dynamic properties [MW]				
		positive	negative	Ntmin	Nt	Ntmax	Ntn	Pn
BLOK1	CoCF	95	-95	899	1349	1899	1913	1900
M4	PoHF	46	-109	40	149	260	234	225
M5	SCoC	0	0	159	174	195	195	195
M6	CoC	0	0	2379	2799	3399	3385	3395
N12	SCoCF	81	-81	1120	1339	1599	1634	1598
N15	SCoCF	143	-143	3920	4564	5600	5721	5593
N16	SCoCF	35	-35	959	1033	1319	1434	1402
BLOK3	CoCF	9	-9	79	174	195	195	195
Tottal		**409**	**-472**	**9555**	**11581**	**14466**	**14711**	**14503**

E. *Primary power control process and dynamics*

BLOK3 switch off time is set to 1 second. Dynamic simulation ends after 40 seconds. The output of dynamic simulation has two charts. First chart is dynamic of frequency deviation in node N15. Chart is shown on figure Fig. 2. Thus the frequency is global parameter the frequency deviation is same in whole power system.

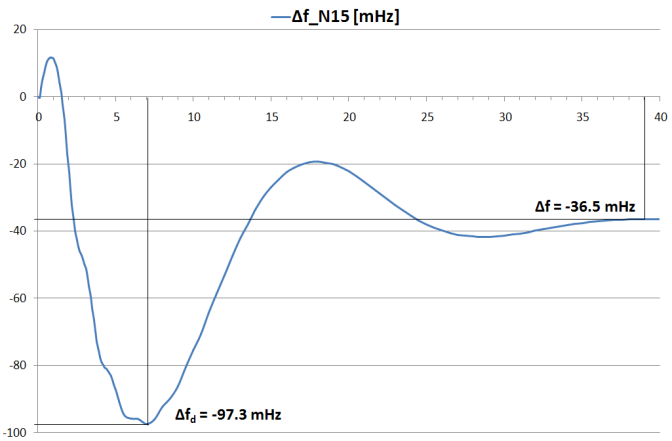


Fig. 2 Frequency deviation - activation of primary power control

On figure 2 is a typical frequency deviation curve after fault and activation of primary control. There is a large decrease of frequency which is shown as increase of frequency deviation

to negative value. This is in process until the primary regulation is fully activated. The minimum of function is called the dynamic frequency divergence marked as Δf_d. Its value is minus 97.3 mHz at time 6 second after a fault. Frequency deviation is stabilized after approx. 30 seconds. This stabilized frequency deviation is called quasistationary frequency divergence marked as Δf. Its value is minus 36.5 mHz

Process of active power activation of individual turbine is shown on figure 3. As it was mentioned before eight generators (turbines) are connected to power grid. Only seven generators are shown because the BLOK3 was disconnected and thus it is not included in chart. Mechanical turbine power is shown in per unit system because of differences in installed power on presented turbo generators.

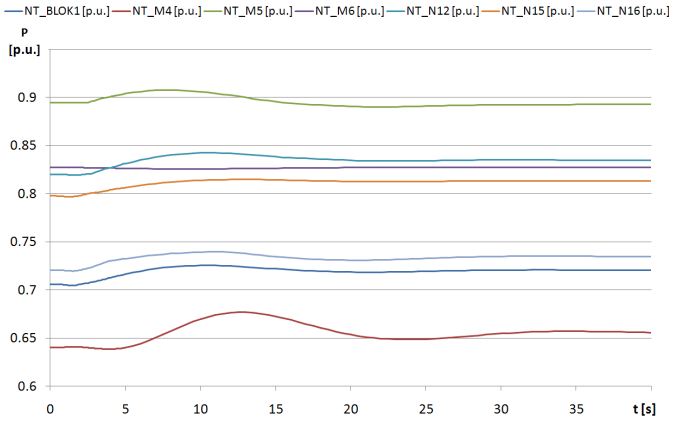


Fig. 3 Primary power control - mechanical power activation on turbines connected to primary regulation process

Red line is for turbine M6. This turbine has a classic type controller without the frequency speed corrector. This turbine maintains the constant mechanical power regardless of frequency deviation in power system.

Turbine M5 has the light blue line and serial configuration of hydraulic speed controller and power controller with no frequency correction. Power activation in this case is not typical for primary regulation. This is due to response of turbine mechanical power to frequency deviation. This controller correct turbine power to preset value after frequency is stabilized.

Green line is associated with turbine M4. Turbine has slower response due to a control mechanism of hydro turbine which has larger dimensions than typical steam valve.

Fastest response has turbine N16 (dark blue line). This turbine has parallel configuration of electronic controller with a frequency corrector and speed controller.

Turbine N12 with purple line has slightly slower response than N16 turbine. Turbine has serial configuration of electronic regulator with frequency corrector and hydraulic speed controller.

Turbine in BLOK1 has also longer response (blue line). This turbine is controlled by electronic controller with frequency correction without speed controller.

Turbines BLOK1, N12 and N16 has primary control reserve set to ±5% of turbine installed power. Hydro turbine doesn't have this limitation thus it can change mechanical power without restriction. Turbine N15 has reserve set to ±1.5% thus its share to primary control process is smaller.

Mechanical power of turbines after primary control activation is for all turbines greater than before activation by the very same value. This is due to same primary regulation static settings (5%).

IV. Frequency Load Shedding

A. Frequency shedding process

In case of more significant frequency deviations a change in turbine control regime (from primary power/frequency control to operation speed control regime) occurs. Frequency load shedding occurs when the primary control reserve is smaller than amount of power which is switched off in power system. This causes the lag of generated power and persistent decrease of system frequency. To stop this decrease and to protect power system against the collapse and blackout the frequency load shedding is activated. This system reduces the load in selected nodes to balance difference between power production and consumption. Dynamic analyses of this process are described in next chapter.

B. Frequency load shedding process and dynamics

Frequency load shedding is a part of node dynamics in MODES. This system works in four independent levels. We can define the frequency load shedding levels for each node. Each level has percentage volume of load reduction related to default steady state. Load reduction level is bounded to exact system frequency. Frequency load shedding settings for this case are in table 3. Frequency levels settings are the same for each node with active frequency load shedding.

TABLE III
FREQUENCY LOAD SHEDDING LEVEL SETTINGS

Name	f_{set1} [Hz]	f_{set2} [Hz]	f_{set3} [Hz]	f_{set4} [Hz]	t_{act1} [s]	t_{act2} [s]	t_{act3} [s]	t_{act4} [s]
UCPTE	49.2	48.7	48.4	48.1	0.1	0.1	0.1	0.1

Table 3 shows that the first level of frequency load shedding starts at system frequency value of 49.2 Hz. The first level is than activated after 100 milliseconds when the system frequency reaches 49.2 Hz. The amount of load shedding in each node with activated frequency load shedding for each level are in table 4.

TABLE IV
AMOUNT OF LOAD SHEDDING IN NODES

Node	Amout of load shedding [%]			
	1st Level	2nd Level	3rd Level	4th Level
N10	17	17	17	17
N12	12	12	12	12
N13	15	15	15	15
N14	12	12	12	12
N15	12	12	12	12
N16	12	12	12	12

To simulate the frequency load shedding the BLOK1 switch off is set in scenario (instead of BLOK3 as it was in previous case). Generator BLOK1 feeds 1350 MW to power system which is more than the primary control reserve (total primary control reserve in power system is 409 MW – table 2) thus first level of frequency load shedding is automatically activated. Figure 4 shows the frequency deviation in node N15 for this case.

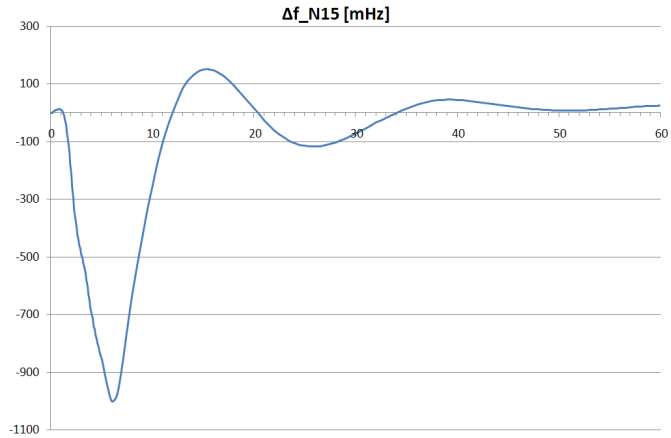


Fig. 4 Frequency deviation – activation of frequency load shedding

When system frequency falls below preset 49.2 Hz the load shedding in selected nodes is activated after the set time interval. This can be seen as a rapid decrease of frequency deviation to value close to zero. System frequency is thus balanced near the nominal system value. This caused amount of summary power of first level of load shedding which is circa 1373 MW (BLOK3 supplies 1350 MW to power system). Frequency deviation after stabilization is 19.9 mHz. Maximum frequency deviation was minus 1000.34 mHz at time circa five second after generator was switched off.

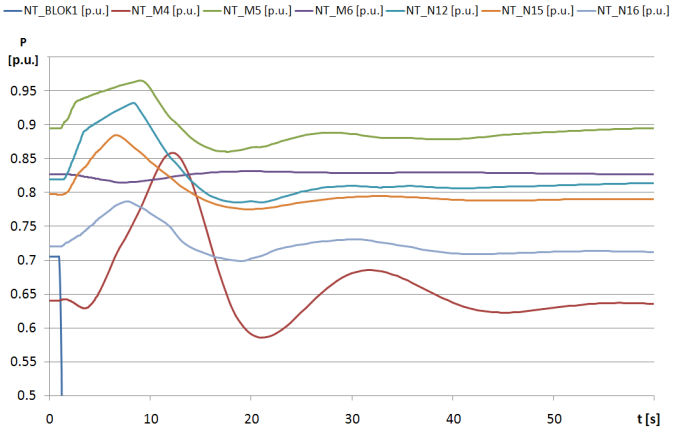


Fig. 5 Frequency load shedding - mechanical power activation on turbines connected to primary regulation process

Mechanical power of turbines in per unit system is shown in figure 5. After the BLOK1was disconnected the turbines start increase mechanical power to cover lack of power in system. After the frequency load shedding is activated and load in nodes is reduced the turbines decrease their mechanical power. Whole process is ensured by primary power control and has analogy with process described in Chapter IV of this paper.

V. Conclusion

This paper described the middle term dynamic in 30 node interconnected power system divided to 4 independent control areas. Activation of primary frequency/power control which is one of the support services used in ENTSO-E the interconnected power system was modeled. This service is automatically activated after the generator disconnection. Primary frequency control monitors the system frequency deviation and activates the reserved power on turbines connected to this service to maintain the power balance. Simulation of this service works like it was described in Chapter III and stabilizes frequency deviation in time of about 30 second. This is the first step in frequency regulation process after fault and lack of generated power in power system. Next step is to decrease the frequency deviation to zero value. This is task for the secondary frequency control which is not a part of analysis described in this paper.

Frequency load shedding is used when the primary control reserve is smaller than amount of disconnected power. The lack of power caused the rapid decrease of system frequency. Decrease bigger than 800 mHz activated the frequency load shedding and reduced the load in selected nodes by set value. Load shedding partially balanced the power between generation and load and with activated primary power control stabilized the frequency deviation. This service also worked as it was described in Chapter IV. This simulation proved that the both described system services can stabilize system frequency and maintain power balance in modeled power system.

This system services was simulated in software like MODES to found out the behavior of generator and power system dynamics. Simulation was done because the measuring or testing activation of power control or frequency load shedding in real power system is not possible (especially frequency load shedding) because it may cause a power system instability, dangerous frequency deviation, even blackout in power system.

Acknowledgment

This work was supported by Slovak Research and Development Agency under the contact No. APVV-0385-07 and VEGA 1/0166/10 and VEGA 1/0388/13 projects.

References

[1] K. Máslo: Střednedobá dynamika, Skripta, Průvodce MODES, 2002.
[2] K. Máslo: Analýza primární regulace, Příručka uživatele, 2001
[3] K. Máslo: Frekvenční odlehčování, Popis modelování, 2001
[4] Z. Trojánek, J. Hájek, P. Kvasnica “Přechodové jevy v elektrizačních soustavách“ 1987 pg. 202–212
[5] K. Máslo „Tvorba dynamických modelů pro praktické výpočty, seminár Aktuální otázky a vybrané problémy řízení ES“. Dostupné na internete www.modesinfo.com
[6] V. Chladný, M Bilička, “ Prechodové javy v elektrizačných sústavách” November 1991
[7] J. Arrillaga & col.: Computer Modelling of Electrical Power System; John Willey & Sons ; 1983
[8] J. Machowski, J.W.Bialek, J.R.Bumby: Power System Dynamics and stability, John Willey & Sons ; 1997
[9] H. Saadat „Power System Analysis“ 2004
[10] M. Kolcun, V. Griger, L. Bena, J. Rusnak:. „Analýza elektrizačnej sústavy” Košice 2005
[11] K. Máslo: Influence of wind farms on transmission system operation in the central Europe, the 9th Interna-tional Conference Control of Power Systems,Tatranské Matliare, Slovakia, 2010,, ISBN 978-80-89409-19-9
[12] Kušnír, S., Beňa, L., Kolcun, M.: The Impact of FACTS Devices to Control the Load Flow. In Proceedings of the 11th International Scientific Conference Electric Power Engineering 2010. Brno: Brno University of Technology, 2010, vol. 1, p. 99-103. ISBN 978-80-214-4094-4.
[13] C. J. Kaufman, Rocky Mountain Research Lab., Boulder, CO, private communication, May 1995.

Assessment of heat source´s heat flux distribution

[1]*Martin BAČKO (4th year)*, [2]*Dmytro MAMCHUR (5th year)*
Supervisor: [3]*Dobroslav KOVÁČ*

[1,3]Dept. of Theoretical Electrical Engineering and Electrical Measurement, FEI TU of Košice, Slovak Republic
[2]Institute of Electromechanics, Energy Saving and Control Systems Kremenchuk Mykhailo Ostrohradskyi National University, Ukraine

[1]martin.backo@tuke.sk, [2]scenter@kdu.edu.ua

***Abstract*—The heat source heat flux distribution assessment is important for proper determination of temperature in room, in its middle, floor and ceiling in order to obtain more precise conduction heat transfer results. In case of more than one storey buildings, only a portion of heat travels downwards (heat by radiation). The amount of this heat needs to be determined for proper results. This paper deals with these problems, first chapter contains the brief introduction to heat fluxes in general, second chapter contains the information about the measured and calculated rooms, third and fourth one is about the determination and explanation of coefficients and the last one is the brief summary with information about the verification measurement, which due to lack of space couldn´t be included.**

***Keywords*—convection, radiation, thermal power, heat flux**

I. Introduction to Heat fluxes

Currently used model for determination of object´s energy requirements uses the average temperature of the room for calculations. This model can be more precise if heat fluxes which create the difference in temperature near the floor and ceiling of the room. As we know, the heat is transferred in three ways – conduction - which occurs mainly in solid state objects, convection – occurs in gases and radiation – which is different in nature from the other two, it has wave character and occurs in every object, which has a temperature.

From these, only two of these ways will be considered – thermal flux by convection of air (Φ_{Rconv}) which due to lower density of warmer air travels only upwards and thermal flux by radiation (Φ_{Rrad}) which travels in every way. For the simplicity sake, we will consider the surface of objects (which are grey opaque objects) as a surface of a blackbody, therefore the radiation from every point of object´s surface will have the same intensity in every direction. (So it will create a half-sphere as seen on Fig.1)

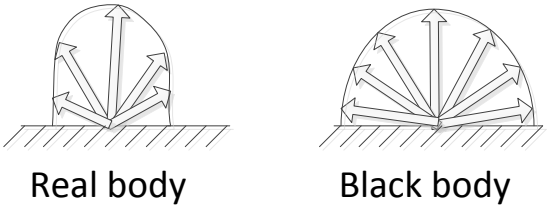


Fig. 1. Real body and black body radiation comparison

If we consider that N points of wall´s or heating device´s surface emits half sphere like the one on Fig.1, then there will be N overlapping half spheres so they can and will be imagined as a plane.

If radiation in form of planes is considered, in case of classic 2 panel radiators we will consider only radiation from the two opposite walls. The radiation from the floor, ceiling and side walls in such abstraction can be considered as insignificant to the radiation of heat source (in this case 2 panel radiation).

Total heat flux of the heat source is:

$$\Phi = \Phi_{Rrad} + \Phi_{Rconv} \quad (1)$$

Thermal heat flux by convection Φ_{Rconv} will be some percentual part of total heat flux Φ.

The determination of approximate percentual distribution of heat flux is essential for more precise temperature near floor and ceiling determination, because then the calculated conductive heat transfer by ceiling and floor will be more precise. The average value of temperature in room can be used in calculation for conductive heat transfer by peripheral walls.

The analysis was made by measurements and calculations of two different rooms and verified on micro model object.

II. Room Measurements and calculations

As mentioned before, two rooms were measured and calculated. They will be labeled as room A and room B. The parameters which will be used for calculations are in following tables (Tab.1 and Tab.2)

Room A parameters:	**Heat source parameters + σ**
a = 4,68 m	a_{rad} = 1,2 m
b = 2,27 m	b_{rad} = 0,6 m
c = 2,53 m	t_{rad} = 40 °C
t_{in} = 21,5 °C	ε_{rad} = 0,8
$t_{ceiling}$ = 22,5 °C	$\sigma = 5{,}68.10^{-8}\,W.m^{-2}.K^{-4}$
t_{floor} = 19 °C	h = 5
ε_{wall} = 0,86	

Tab.1 Room A parameters

ε_{wall} is the emissivity coefficient of concrete wall with common type of limecement plaster, ε_{rad} is the emissivity coefficient of cold rolled steel, from which the radiator is made. σ is the Stefan-Boltzmann´s constant, h is the parameter

of natural free air convection. This coefficient ranges from 5-25, and unfortunately is not explained well, so this is the weakest part of the calculation, with the most probability of error. Because the windows and doors were closed and there were no other sources which could cause and force convection of air, the value of coefficient was considered as minimal, therefore 5.
Table 2 (Tab.2) shows the parameters of second room, room B

Room B parameters:	**Heat source parameters + σ**
a = 3,33 m	a_{rad} = 1,2 m
b = 3,02 m	b_{rad} = 0,6 m
c = 2,53 m	t_{rad} = 33 °C
t_{in} = 22 °C	ε_{rad} = 0,8
$t_{ceiling}$ = 23 °C	$\sigma = 5{,}68.10^{-8}$ W.m^{-2}.K^{-4}
t_{floor} = 19,5 °C	h = 5
ε_{wall} = 0,86	

Tab.2 Room B parameters

Heat flux by radiation from heat source (radiator) Φ_{Rrad}:

$$\Phi_{Rrad} = \sigma * \varepsilon_{rad} * T_{rad}^4 * S \quad (2)$$

Trad is the absolute value of radiator (273,15+40) K. Its surface area is:

$$S = 4 * a_{rad} * b_{rad} \quad (3)$$

because the radiator consist of 2 plates, each of them has 2 sides which radiate and where the heat from surface to surrounding air is transferred.

$$\Phi_{Rrad} = 5{,}68{,}10^{-8}*0{,}8*313{,}15^4*4*1{,}2*0{,}6 = 1258{,}46 \text{ W} \quad (4)$$

Heat flux of heat source by convection Φ_{Rconv} is:

$$\Phi_{Rconv} = k * (T_{rad} - T_{in}) * S \quad (5)$$

$$\Phi_{Rconv} = 5 * (313{,}15 - 294{,}65) * 2{,}88 = 266{,}4 \text{ W} \quad (6)$$

As can be seen, the radiative heat flux from radiator is several times higher as the convective heat flux. However we have to consider the radiation from the opposite wall, which will affect the radiation from heat source.
It will calculated analogically:

$$\Phi_{wall1} = \sigma * \varepsilon_{wall} * T_{wall1}^4 * S_{wall1} \quad (7)$$

$$\Phi_{wall1} = 5{,}68*10^{-8}*0{,}86*(273{,}15+21{,}5)^4*2{,}27*2{,}53 = 2114{,}55\text{W} \quad (8)$$

The wall behind the radiator will be calculated analogically, the only difference will be in area considered, which will equal the area of heat source, because only this portion of surface will actively radiate against the heat source.

$$\Phi_{wall2} = 5{,}68*10^{-8}*0{,}86*(273{,}15+21{,}5)^4*1{,}2*0{,}6 = 265{,}09\text{W} \quad (9)$$

The difference between heat fluxes of wall and heat sources:

$$\Delta\Phi_1 = \Phi_{Rad} - \Phi_{wall1} = 1258{,}46 - 2114{,}55 = -856{,}09 \text{ W} \quad (10)$$
$$\Delta\Phi_2 = \Phi_{Rad} - \Phi_{wall2} = 1258{,}46 - 265{,}09 = 993{,}36 \text{ W} \quad (11)$$

As can be seen from $\Delta\Phi 1$ value, the radiative heat flux from wall 1 is higher than flux from heat source (negative delta value), which would mean that the wall is more effective heat source than radiator, which of course is nonsense. Both of the fluxes will have to multiplied by coefficients (k_1 and k_2) which will take into account the area and distance of the wall from the heat source.

For the next calculation, following assumption will be used: Convective heat flux Φ_{Rconv} is a part of total heat flux Φ. The 5% steps were used. Following equation shows an example of calculation in case that Φ_{Rconv} is 40% of total Φ.

$$\Phi_{40} = (\Phi_{Rconv}/40)*100 = (266{,}4/40)*100 = 666\text{W} \quad (12)$$

Following table (Tab.3) shows the value for each percentual step with coefficient k1 which will represent the distance and area of the opposite wall.

$$\Delta\Phi_1*k_1 + \Delta\Phi_2*k_2 = \Phi \quad (13)$$

The area of wall behind the heat source (wall 2) will always be the same like the area of heat source and its distance will always be the same, therefore the thermal exchange will occur between the same areas in constant distance. Because of this, the coefficient k_2 value will be set to value 1.

Room A			
If	**Expected total Φ**	**k1**	**kx = k1/a**
ΦRconv is 10%Φ	2664	-1,95146618	-0,41698
ΦRconv is 15%Φ	1776	-0,914194176	-0,19534
ΦRconv is 20%Φ	1332	-0,395558173	-0,08452
ΦRconv is 25%Φ	1065,6	-0,084376572	-0,01803
ΦRconv is 30%Φ	888	0,123077829	0,026299
ΦRconv is 35%Φ	761,1428571	0,271259544	0,057961
ΦRconv is 40%Φ	666	0,38239583	0,081709
ΦRconv is 45%Φ	592	0,468835164	0,100178
ΦRconv is 50%Φ	532,8	0,537986631	0,114954
ΦRconv is 55%Φ	484,3636364	0,594565104	0,127044
ΦRconv is 60%Φ	444	0,641713831	0,137118
ΦRconv is 65%Φ	409,8461538	0,681608908	0,145643
ΦRconv is 70%Φ	380,5714286	0,715804689	0,15295
ΦRconv is 75%Φ	355,2	0,745441032	0,159282
ΦRconv is 80%Φ	333	0,771372832	0,164823
ΦRconv is 85%Φ	313,4117647	0,794253832	0,169712
ΦRconv is 90%Φ	296	0,814592499	0,174058
ΦRconv is 95%Φ	280,4210526	0,832790253	0,177947
ΦRconv is 100%Φ	266,4	0,849168232	0,181446

Tab.3 Room A calculations

Column k_1/a represents the value of coefficient regarding the unit of distance, in this case 1 meter. The new coefficient is labeled as k_x:

$$k_x = k_1/a \quad (14)$$

where a = 4,68 which is the length of the room and distance of the wall from the heat source.

This k_x coefficient is only valid for Room A, therefore, another measurement and calculation will have to be made (Room B, Tab.2).

Calculations for room B are analogic:
Heat flux by radiation from heat source (radiator):

$$\Phi_{Rrad} = 5{,}68{,}10^{-8}*0{,}8*306{,}15^4*4*1{,}2*0{,}6 = 1149{,}65\ W \quad (15)$$

Heat flux by convection Φ_{Rconv} is:

$$\Phi_{Rconv} = 5 * (306{,}15 - 295{,}15) * 2{,}88 = 158{,}4\ W \quad (16)$$

Radiative heat flux from opposite wall (wall 1):

$$\Phi_{wall1} = 5{,}68*10^{-8}*0{,}86*(273{,}15+22)^4*3{,}02*2{,}53 = 2832{,}34W \quad (17)$$

Radiative heat flux from wall behind the radiator (wall 2):

$$\Phi_{wall2} = 5{,}68*10^{-8}*0{,}86*(273{,}15+22)^4*1{,}2*0{,}6 = 266{,}09W \quad (18)$$

The difference between heat fluxes of wall and heat sources:

$$\Delta\Phi_1 = \Phi_{Rad} - \Phi_{wall1} = 1149{,}65 - 2832{,}32 = -1682{,}68\ W \quad (19)$$
$$\Delta\Phi_2 = \Phi_{Rad} - \Phi_{wall2} = 1149{,}65 - 266{,}09 = 882{,}75\ W \quad (20)$$

Calculated values for room B are shown in following table (Tab.4)

Room B			
If	**Expected total Φ**	**k1**	**kx = k1/a**
ΦRconv is 10%Φ	1584	-0,41674	-0,125147761
ΦRconv is 15%Φ	1056	-0,10296	-0,030918222
ΦRconv is 20%Φ	792	0,053935	0,016196548
ΦRconv is 25%Φ	633,6	0,14807	0,044465409
ΦRconv is 30%Φ	528	0,210827	0,063311317
ΦRconv is 35%Φ	452,5714	0,255653	0,07677268
ΦRconv is 40%Φ	396	0,289273	0,086868702
ΦRconv is 45%Φ	352	0,315421	0,094721164
ΦRconv is 50%Φ	316,8	0,33634	0,101003133
ΦRconv is 55%Φ	288	0,353456	0,106142926
ΦRconv is 60%Φ	264	0,367719	0,110426087
ΦRconv is 65%Φ	243,6923	0,379787	0,1140503
ΦRconv is 70%Φ	226,2857	0,390132	0,117156768
ΦRconv is 75%Φ	211,2	0,399097	0,119849041
ΦRconv is 80%Φ	198	0,406942	0,122204779
ΦRconv is 85%Φ	186,3529	0,413864	0,124283372
ΦRconv is 90%Φ	176	0,420016	0,12613101
ΦRconv is 95%Φ	166,7368	0,425521	0,12778416
ΦRconv is 100%Φ	158,4	0,430476	0,129271995

Tab.4 Room B calculations

As before:

$$kx = k1/a \quad (21)$$

where a = 3,33 which is the length of the room.
The value of the k_x coefficients needs to compared. The row where the difference between k_x coefficients will be the lowest, this will be the most probable percentual distribution.

III. Determination of K_X coefficient

Following table (Tab.5) shows the coefficients k_x comparison. Third column represents the difference between them. The difference is calculated in absolute value. As can be seen, the lowest difference (the most precise guess) ranges when Φ_{Rconv} ranges between 40-45% of total Φ. This interval will be divided into steps of 1% in order to obtain more precise result.

ΦRconv from Φ	**Room A kx**	**Room B kx**	**Absolute value of difference**
10%	-0,416979953	-0,125147761	0,291832192
15%	-0,195340636	-0,030918222	0,164422414
20%	-0,084520977	0,016196548	0,100717525
25%	-0,018029182	0,044465409	0,062494592
30%	0,026298681	0,063311317	0,037012636
35%	0,057961441	0,07677268	0,018811239
40%	0,081708511	0,086868702	0,005160191
45%	0,100178454	0,094721164	0,00545729
50%	0,114954408	0,101003133	0,013951275
55%	0,127043826	0,106142926	0,0209009
60%	0,13711834	0,110426087	0,026692253
65%	0,145642929	0,1140503	0,031592629
70%	0,15294972	0,117156768	0,035792952
75%	0,159282272	0,119849041	0,039433231
80%	0,164823255	0,122204779	0,042618475
85%	0,169712357	0,124283372	0,045428985
90%	0,174058226	0,12613101	0,047927216
95%	0,177946635	0,12778416	0,050162475
100%	0,181446203	0,129271995	0,052174209

Tab.5 Comparison of k_x coeficient

Following table (Tab.6) shows the calculations for room A, if we divide the interval from 40% to 45% by 1% step.

If	**Expected total Φ**	**k1**	**kx = k1/a**
ΦRconv is 40%Φ	666	0,38239583	0,081709
ΦRconv is 41%Φ	649,7560976	0,401370318	0,085763
ΦRconv is 42%Φ	634,2857143	0,419441259	0,089624
ΦRconv is 43%Φ	619,5348837	0,436671691	0,093306
ΦRconv is 44%Φ	605,4545455	0,453118921	0,09682
ΦRconv is 45%Φ	592	0,468835164	0,100178

Tab.6 1% steps calculation room A

The same is done for room B (Tab.7)

If	**Expected total Φ**	**k1**	**kx = k1/a**
ΦRconv is 40%Φ	396	0,289273	0,086868702
ΦRconv is 41%Φ	386,3415	0,295013	0,088592413
ΦRconv is 42%Φ	377,1429	0,300479	0,090234043
ΦRconv is 43%Φ	368,3721	0,305692	0,091799318
ΦRconv is 44%Φ	360	0,310667	0,093293443
ΦRconv is 45%Φ	352	0,315421	0,094721164

Tab.7 1% steps calculation room B

Final table (Tab.8) shows the comparison of these two k_x coefficients:

If	kxa = k1/a	kxb = k1/a	difference
ΦRconv is 40%Φ	0,086868702	0,081708511	0,005160191
ΦRconv is 41%Φ	0,088592413	0,085762888	0,002829525
ΦRconv is 42%Φ	0,090234043	0,089624201	0,000609842
ΦRconv is 43%Φ	0,091799318	0,093305917	-0,001506599
ΦRconv is 44%Φ	0,093293443	0,096820282	-0,003526839
ΦRconv is 45%Φ	0,094721164	0,100178454	-0,00545729

Tab.8 Comparison of the k_x coefficients of room A and room B

The lowest difference in absolute value is highlighted in yellow in Table 8. As can be seen, it is when convective heat flux Φ_{Rconv} is 42% of total heat flux Φ. This value can be considered as a percentual distribution which can be used for following calculations.

The final value of k_x coefficient will be set as an average value of the k_{xA} and k_{xB} coefficients for 42% distribution.

$$k_x = (0{,}090234043+0{,}089624201)/2 = 0{,}0899 \quad (22)$$

The general equation for total thermal heat flux generated by heat source will be:

$$(\Phi_{Rconv} - \Phi_{wall1}) * 0{,}0899 * a + (\Phi_{Rconv} - \Phi_{wall2})*1 = \Phi \quad (23)$$

where a is the distance of wall 1 (opposite one) from heat source.

IV. Determination of Heat Gradient Coefficient

As can be seen from the rooms measurement, the temperature near the ceiling is not the same as the temperature in the middle of the room (where the thermometer is placed usually). Roughly the differences are about 1°C increase near the ceiling and 1°C decrease near the floor. These differences are created because of the heat fluxes Φ_{Rrad} and Φ_{Rconv}.

Considering the STN 73 4301 which speaks about the residential buildings the standard height of the ceiling is between 2,4 – 2,6 m. Older buildings can have the ceilings higher, the thermal gradient has to be for the unit of distance (1 meter).

Heat gradient for room A:

$$\Delta t_a = t_{ceiling} - t_{floor} = 22{,}5\text{-}19 = 3{,}5°C \quad (24)$$

For room B:

$$\Delta t_b = t_{ceiling} - t_{floor} = 23\text{-}19{,}5 = 3{,}5°C \quad (25)$$

In both cases, the gradient has the same value, both rooms have the same height (2,53m).

For the unit of distance the gradient t_{grad} will be:

$$t_{grad} = \Delta t /2{,}53 = 3{,}5/2{,}53 = 1{,}38 \quad (26)$$

For the length of 1 meter the change is 1,38 °C.

This coefficient will be used in calculations, where the height of the room specified by user will be divided by 2, multiplied by t_{grad} coefficient and added to the temperature of room desired by user. (This will represent the temperature increase near the ceiling.).

The same will be done for room near the floor, except that the result from t_{grad} coefficient will be substituted from the temperature of the room specified by user. (This will represent the temperature decrease near the floor).

V. Conclusion

Calculated coefficients provide us with more precise model for determining the thermal power, which is needed for the room. These calculations were verified with the experiment on micro model where based upon the desirable temperatures in rooms, the thermal power needed was calculated and then measured. The differences between calculated and measured values were only about 10% therefore, we consider the obtained results acceptable for the model. Unfortunately, due to lack of space, the verification measurement and experiment cannot be included and will be a subject of different paper.

Acknowledgment

The paper has been prepared under support of Slovak grant project KEGA No. 005TUKE-4/2012.

References

[1] Cengel Yunus A: Introduction to Thermodynamics and Heat transfer, McGraw Hill, 2008 England

[2] Tritt Terry, M: Thermal Conductivity : Theory, properties and application, Kluver Academics, USA

[3] Kovac Dobroslav, Kovacova Irena: Non-harmonic Power Measuring, Acta Electrotechnica et Informatica 3/2008

[4] Methodology for meeting object´s energy demands by renewable energy sources / Martin Bačko, Dobroslav Kováč, Julia Alekseeva - 2011. In: Electromechanical and energy saving systems. No. 2 (2011), p. 81-85. - ISSN 2072-2052 [BAČKO, Martin - KOVÁČ, Dobroslav - ALEKSEEVA, Julia]

[5] Program for determinig the most economical solution for object´s heating purposes / Martin Bačko, Dobroslav Kováč - 2012. In: Electromechanical and energy systems, modeling and optimization methods : conference proceedings : the 10th international conference of students and young researchers : March 28-29, 2012, Kremenchuk. - Kremenchuk : Kremenchuk Mykhaylo Ostrogradskiy National University, 2012 P. 83-84. - ISSN 2079-5106 [BAČKO, Martin - KOVÁČ, Dobroslav]

Comparison of the dissipation factor and the relative permittivity of mineral oils and natural esters

[1]*Lukáš Lisoň (1st year),* [2]*Marián Hrinko(3rd year)*
Supervisor: [3]*Iraida Kolcunová*

[1]Dept. of Electrical power engineering, FEI TU of Košice, Slovak Republic
[2,3]Dept. of Electrical power engineering, FEI TU of Košice, Slovak Republic

[1]lukas.lison@tuke.sk, [2]marian.hrinko@tuke.sk, [3]iraida.kolcunova@tuke.sk

***Abstract*— The article is focused on the measuring of the dissipation factor and the relative permittivity of mineral oils and natural esters used in electric power devices. Natural esters can be used as refund classical mineral oils. This work is denote on measurement relative permittivity and dissipation factor in depending on voltage.**

***Keywords*— relative permittivity, dissipation factor, mineral oil, natural ester**

I. Introduction

Insulation system is the weakest part of each electric power device. This part is devoted a lot of attention, because good choice of insulation can extend lifetime of device and save a lot of money. The insulation system of most transformers operating in Slovakia used combination solid and liquid insulation. For most transformers, oils are used as impregnation and cooling medium. Mineral oil is obtained from petroleum as one fraction in the distillation process [1]. Natural esters are dielectrics, which are made from renewable plants crops [2]. Compared to mineral oils, natural esters are more environmentally friendly and it is very important condition today

II. Dissipation factor

Dissipation factor expresses degree of dielectric loss. In case of measuring of the dissipation factor, must be insulation connected like dielectric of capacitor. Phase shift between current and voltage of ideal dielectric is 90°. Although, insulation materials are not ideal, current flow across dielectric has except reactive and also real part so phase shift is smaller than 90° [2].

Dissipation factor, denoted as tan δ is defined as tangent of the angle about which differ phase shift of real dielectric against phase shift current of ideal dielectric. Tan δ is connected to dielectric loss and values of the dissipation factor it can considered as insulation quality scale. On fig.1 is shown phasor diagram of ideal dielectric with only one polarizing storyline [3].

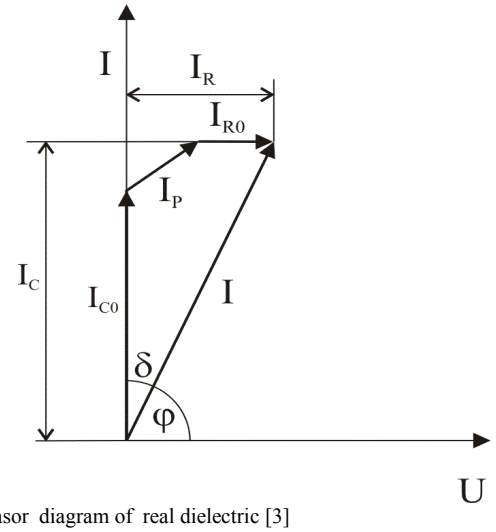


Fig. 1. Phasor diagram of real dielectric [3]

Where : U – applied alternating voltage
I – current flowing across dielectric
I_C – capacity part of current of real dielectric
I_R – real part of current of real dielectric
I_P – polarization current
I_{RO} – conduction current

Tan δ is defined as ratio real part and capacitive part of current .

$$tg\delta = \frac{I_R}{I_C} \tag{1}$$

Dissipation factor is a dimensionless value. Dielectric losses were measured by using Schering bridge. Scheme of Schering bridge is on fig.2 where test subject C_X and capacitive normal C_N create a high voltage side of bridge and variable resistors and capacitors which are used to balance of the bridge, create low voltage side of bridge.

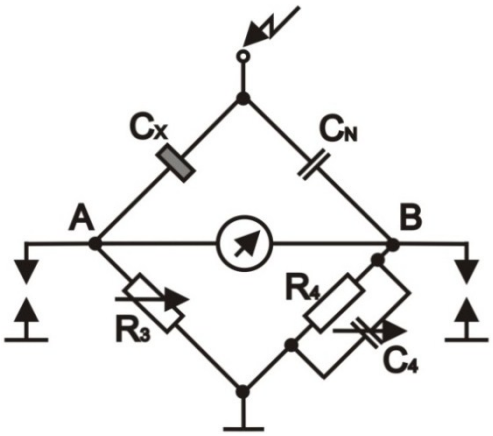


Fig. 2. Schering bridge [3]

Tan δ can be defined from condition of balance the bridge as:

$$tg\delta = \omega C_4 R_4 \qquad (2)$$

III. Relative permittivity

Relative permittivity ε_r it can defined as ratio of capacity dielectric between capacitors electrodes and capacity of the same capacitor which is filled by air. It describes how many times is force exerted on charge in dielectric lower than in vacuum [5].

$$\varepsilon_r = \frac{C_X}{C_O} \qquad (3)$$

Relative permittivity is used to expresses of skills of material accumulate electric charge. Accumulation of electric charge is consequence of polarization of matarial. Polarization is moving of a free charges in electric field. It depend of temperature of specimens and frequency of voltage. In nonpolar liquid insulators occures only electron polarization. Polar liquid insulators have an electron and dipole polarization. Relative permittivity is depends on dipole moments of molecules, and their speed in electrical field[6].

The permittivity of many substances is changed not only with frequency and temperature, but also with the age of specimens. That means that two of the same specimens of liquid insulators can have the different value of relative permittivity. It depends on conditions, where the specimens were used.

IV. Measured specimens and measurement process

Measuring of relative permittivity and the dissipation factor were realized by using automatic Schering bridge with container on liquid insulators fig. 3. The values of relative permittivity were calculated by equation (3).

Fig.3. Container on measurement Cx and tan δ of liquid insulators [2].

Volume of container is 40 ml, distance of electrodes is 2 mm and capacity C_0 is 60 pF. Measurement was realized at temperature 20°C and voltage was increased from 0.2 kV to 2 kV with step 0.2 kV at frequency 50 Hz. Measurement workplace is shown on fig. 4.

Fig.4 Measurement workplace on measuring C_x and tan δ [2].

Container was cleaned and parched before the measurement. Specimens were poured to container six hour before measuring in order to stabilize oils after the pouring. e.g. removing air bubbles. This process were repeated before each of change of specimens. Values of tan δ and C_x for each voltage were listed on display.

V. Specimens

This specimens were used at measurement tan δ and ε_r of mineral oils and natural esters.

- sunflower oil filtered and chemical adapted to using in food industry (**RS**)
- rapeseed oil filtered and chemical adapted to using in food industry (**RR**)
- inhibited transformer oil ITO 100 highly refined mineral oil with additive 2,6-di-terc-butyl-metylfenolu (**MI**)
- inhibited transformer oil Nynas - Lyra X (**ML**)

VI. RESULTS AND DISCUSSION

Tab. 1 shown the results of measurement of the dissipation factor for various specimens of liquid insulators in depend on voltage.

TABLE I
MEASURED VALUES OF TANGENT DELTA

U[kV]	tan δ[-]			
	RR	RS	MI	ML
0.2	0.0016	0.0048	0.0005	0.0007
0.4	0.0015	0.0052	0.0006	0.0008
0.6	0.0014	0.0046	0.0003	0.0004
0.8	0.0014	0.0051	0.0003	0.0003
1	0.0015	0.0052	0.0003	0.0002
1.2	0.0015	0.0072	0.0002	0.0002
1.4	0.0015	0.0123	0.0002	0.0016
1.6	0.0015	0.0112	0.0002	0.0107
1.8	0.0016	0.0088	0.0002	0.024
2	0.0019	0.0085	0.0001	0.0112

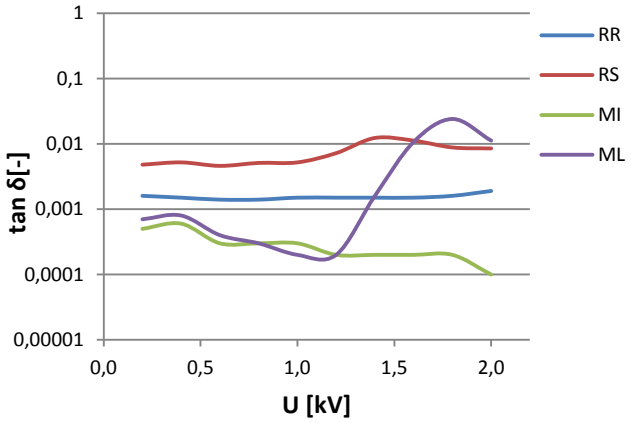


Fig.5 Tan δ in depend on voltage

On fig. 5 it can see tan δ in depend on voltage. Both natural esters achieves higher value than mineral oils. The highest value was measured of the sunflower oil. Mineral oils had the same waveforms to the voltage 1 kV. Above the value 1kV tan δ of mineral oil ML rise and reach higher values than natural esters.

TABLE 2
MEASURED VALUES OF RELATIVE PERMITTIVITY

U[kV]	ε_r[-]			
	RR	RS	MI	ML
0.2	3.255	3.330	2.275	2.310
0.4	3.253	3.328	2.275	2.310
0.6	3.253	3.327	2.275	2.310
0.8	3.253	3.327	2.277	2.310
1	3.253	3.327	2.277	2.310
1.2	3.253	3.327	2.277	2.308
1.4	3.253	3.327	2.277	2.310
1.6	3.253	3.327	2.277	2.305
1.8	3.253	3.327	2.277	2.303
2	3.253	3.328	2.277	2.308

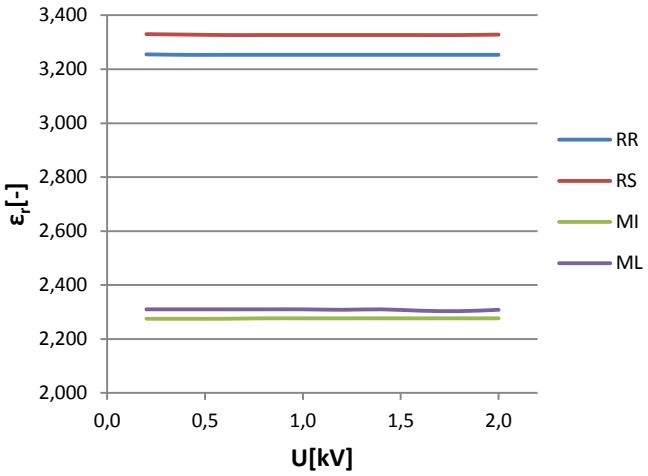


Fig.6 Relative permittivity in depend on voltage

The results of measuring of relative permittivity shown that natural esters have higher value than mineral oils. On fig. 6 it can see that values of relative permittivity were unchanging in depend on voltage for each type of oils. Relative permittivity of natural ester was in ranging from 3.25 to 3.33, whilst the mineral oils have the values in ranging from 2.27 to 2.31.

VII. CONCLUSION

This article was focused on measurement of the dissipation factor and the relative permittivity natural esters and mineral oils, which can be used as substitute of conventional mineral oils, mainly from environmental aspect. Dissipation factor as one of factors pointing at quality of insulation system, reaches higher values for natural ester. That means that natural esters have a higher dielectric losses as a mineral oils. Relative permittivity was markedly unchanged for the voltage ranging from 0.2 to 2 kV. Measurement of the dissipation factor and the relative permittivity is not sufficient to consider of changes mineral oils to natural esters. Next research should be focused on measuring of others properties those liquid insulators.

ACKNOWLEDGMENT

We support research activities in Slovakia / Project is co-financed from EU funds. This paper was developed within the Project "Centrum excelentnosti integrovaného výskumu a využitia progresívnych materiálov a technológií v oblasti automobilovej elektroniky", ITMS 26220120055

The authors also wish to acknowledge Scientific Grant Agency of the Ministry of Slovak Republic and Slovak Academy of Science for funding of experimental works in the frame of VEGA No. 1/0487/12 grant.

REFERENCES

[1] ARTBAUER, ŠEDOVIČ, ADAMEC : Izolanty a izolácie ,ALFA, Bratislava ,1969

[2] DEDINSKÁ : Vplyv elektrotepelného namáhania na elektroizolačné vlastnosti rastlinných olejov, Dizertačná práca , Košice 2012.

[3] KOLCUNOVÁ,I.: Diagnostiky v elektroenergetike, Prednašky pre 5.roč KEE Košice,2011, Dostupné na internete: http://web.tuke.sk/fei-kee/web/index.php?pg=diagnostika-v-elektroenergetike&hl=sk

[4] KOLCUNOVÁ.I.: Diagnostika elektrických strojov. Technická univerzita Košice 2006 s. 22 -23 ISBN 80-8073-550-6

[5] MARCI : Výskum výbojových procesoch v kvapalných dielektrikách, Dizertačná práca , Košice 2012

[6] DEDINSKÁ, L – KOLCUNOVÁ,I: Elektrické vlastnosti prírodných esterov. Publikácie TUKE Starnutie elektroizolačných systémov , č9 (2010) ISSN 1337-0103

Comparison of least squares and maximum likelihood fitting for ADC testing

[1]*Jozef LIPTÁK (2nd year)*, [2]*Marek GODLA (3nd year)*
Supervisor: [3]*Ján ŠALIGA*

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

[1]jozef.liptak@tuke.sk, [2]marek.godla@tuke.sk, [3]jan.saliga@tuke.sk

***Abstract*—When we make a dynamic ADC test, we need to reconstruct the input signal as best as possible from the acquired ADC output data. This paper presents an experimental comparative study of estimation of the signal parameters using well known least squares and recently introduced maximum likelihood method. In the conclusion are some practical recommendations for the choice of optimal method for various conditions of ADC.**

***Keywords*—ADC testing, least squares estimation, maximum likelihood method, sine fitting, signal reconstruction, differential evolution.**

I. Introduction

Standardized dynamic test methods for analog-to-digital converters (ADCs) ([1] – [3]) are based on comparison of acquired ADC output codes with the ADC input stimulus. The stimulus is not exactly known therefore it must be reconstructed from the erroneous ADC output codes acquired during testing. Any inaccuracy in the estimation of ADC stimulus parameters leads to inaccuracy in determination of ADC parameters measured by the dynamic test.

According to the theory the least squares (LS) fitting gives the best estimation under the condition that the observation (quantization) noise is additive to the input, is independent, white, and normally distributed with zero mean ([4], [5]). This all is clearly not true for ADC testing and therefore the LS fit is usually worse than ML estimation would be.

The general, systematic "best" way of estimation is fitting based on the maximum likelihood (ML) method. This idea was introduced in [6] for sine wave and later it was generalized also for exponential stimulus ([7]).

The main novelty of this paper is just a comparative research. Moreover, we also examined the differential evolution based optimization method for ML fit [9]. We performed calculations and signal processing in LabVIEW.

II. Fitting methods

The general setup for dynamic ADC testing is shown in Fig. 1. To perform simulation experiments we developed a few software modules including non-ideal ADC model with optional test specific INL. The modules enable simulating real ADC test according to Fig. 1.

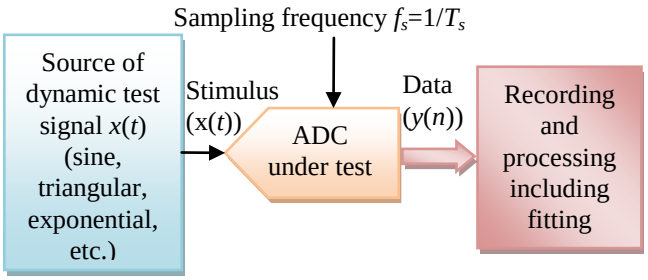


Fig. 1. General setup for dynamic ADC testing

A. Least Squares (LS) fit

LS fitting is a very often used procedure for recovering distorted and noisy signals in testing and measurements. Estimation of parameters of ADC input signal is obtained from minimization of an cost function. Although this method is well known we will not pay any further attention to it.

B. Maximum likelihood (ML) fit

Maximum likelihood method comes from estimation theory and looks for the most probable ADC input signal. The ML fitting is based on maximization of a likelihood function $L(\mathbf{a})$:

$$\max_{\mathbf{a},\mathbf{q},\sigma}\left(L(\mathbf{a})\right)=\max_{\mathbf{a},\mathbf{q},\sigma}\prod_{n=0}^{N-1}P\left(y(n)=Y_{\mathbf{a},\mathbf{q},\sigma}(n)\right), \tag{1}$$

where $\mathbf{q}$ is the vector of ADC transition code levels, σ is the standard deviation of the Gaussian noise of the input electronics, assuming here that the noise samples are independent. $P(y(n) = Y(n))$ is the probability that the n-th ADC output sample $y(n)$ is equal to the expected output value $Y(n)$, which represents a possible output code. To simplify the maximization $L(\mathbf{a})$, minimization of the sum of negative logarithms of probabilities $P(.)$ is preferred.

$$\begin{aligned}\arg\max_{\mathbf{a},\mathbf{q},\sigma}\left(L(\mathbf{a})\right)&\approx\arg\min_{\mathbf{a},\mathbf{q},\sigma}\left(-\ln\left(L(\mathbf{a})\right)\right)=\\&=\arg\min_{\mathbf{a},\mathbf{q},\sigma}\left(-\sum_{n=0}^{N-1}\ln P\left(y(n)=Y_{\mathbf{a},\mathbf{q},\sigma}(n)\right)\right)\end{aligned} \tag{2}$$

Searching the extreme value can be a complex task and it can be performed only by an appropriate numerical method.

C. *Implementation of fitting and comparison*

Experiments below were primarily performed in software developed in LabVIEW. Properties, advantages and disadvantages of LabVIEW are generally well known. We choose LabVIEW not only because of our previous experience but also because we wanted to compare our results with previous experiments performed in Matlab ([6], [7]).

LS fit for a sine wave was implemented according to standards ([1] – [3]) in the form of 3- and 4-parameter fits. ML fit implementation was rather complex. In contrast to minimization methods based on gradients used in [6] and [7], the LabVIEW built-in function `GlobalOptimization.vi` ([8]) has been utilized. This function is based on the Differential Evolution method (DE), which is a kind of genetic algorithm, being one of the general methods used to solve the global optimization problem [9].

For comparison of accuracy of signal recovering achieved by LS and LM methods we utilized two parameters:

1. Estimation error defined as RMS value of the difference between estimated signal and the exact input signal.
2. Difference between SINADs calculated from LS fit, or from ML fit and the exact value calculated from exactly known ADC stimulus.

The RMS value was chosen because it is the basic descriptor of the error which has influence on nearly all final ADC dynamic parameters, such as SINAD and ENOB. The estimation error D_{method} (RMS value of the difference between the recovered signal by a given method and the known exact one) was calculated according to the formula:

$$D_{\text{method}} = \sqrt{\frac{1}{N}\sum_{n=1}^{N-1}\left(y_{\text{method}}(n) - y_{\text{exact}}(n)\right)^2}\,, \qquad (3)$$

where $y_{\text{method}}(n)$ are the samples calculated from signal recovered by the given method and $y_{\text{exact}}(n)$ are the equivalent samples of the exact ADC stimulus.

Comparison of SINAD calculated from LS and ML fits to SINAD calculated from the exact value of stimulus was utilized only in tests, where the stimulating sine wave did not overload ADC full scale as it is required by standards. SINAD in dB was calculated according to the standard defined formula:

$$SINAD_{\text{method}} = 20\log\frac{A_{\text{method}}}{D_{\text{method}}}\,, \qquad (4)$$

where A_{method} is RMS value of sine wave recovered by a given method or from exactly known ADC stimulus.

III. Simulation Results

The first experiments were performed on linear ideal 8-bit ADC and standard deviation of noise equal to 0.2 LSB, which represented ADC code alternating behaviour.

When we compare in time domain the exact stimulus that is in the full scale range of the ADC input, and the recovered signals using LS and ML fit, the differences between recovered signals and exact stimulus are hardly visible.

As soon as the input signal slightly overloads the 8-bit ADC full scale range, the differences between the recovered signals became visible, see Fig. 2. For both fits all samples in record were used including samples taken in time instances when signal overloaded ADC input range (overloading samples). In addition the same pre-processed record with the overloading samples excluded was fitted by LS method. In accordance with expectations, ML fit gives much better estimation of ADC stimulus than LS fit of the whole record because of its nature to approach also the extreme ADC overloading samples. If LS fit is applied on a pre-processed record where overloading samples are excluded from calculation the results obtained by LS and ML are nearly the same (Fig. 2).

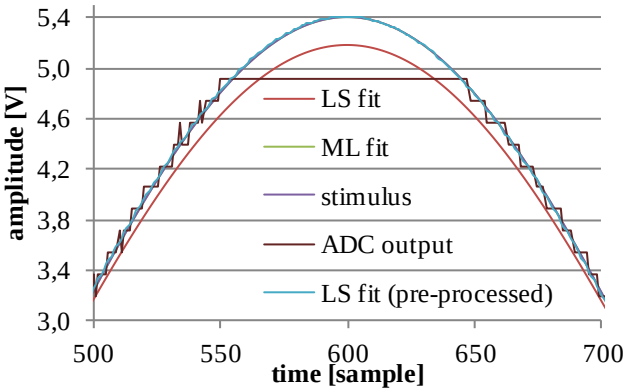


Fig. 2. Comparison of LS and ML fit, digitized and original stimulus (zoomed time segment) if stimulus overloads full range of the ADC. ML fit utilized all samples in record, LS fit utilized once all samples and next the record with excluded overloading samples.

The dependence acquired for an 8-bit ADC and 500 coherently taken samples with standard deviation of noise σ = 0.2 LSB for varying ADC overloading was as expected: if the input signal covers ADC range up to 100 percent, both methods give nearly the same results. RMS value of the difference of LS fit from exact stimulus quickly increases if input stimulus even only slightly overloads ADC input range and if LS fit is applied on whole record. Very similar results were achieved for other ADC resolutions, σ, and numbers of samples in record.

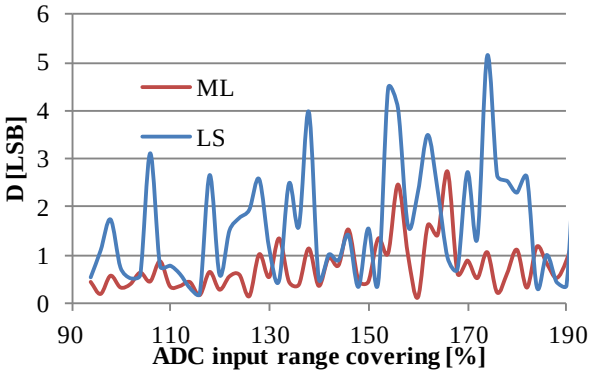


Fig. 3. LS vs. ML fit for overloaded ADC (100% is equal to ADC full scale). LS fit uses pre-processed record.

Fig. 3 shows the same circumstances but LS fit was applied on pre-processed records with excluded overloading samples. At these circumstances, the estimation errors for both fits are small and comparable. The only difference between fits is that the ML fit gave more "stable" error what indicates its smaller sensitivity to quantization effects.

The resulting practical conclusions coming from the above performed experiments is that

- the dominating error factor in "blind" LS fitting is ADC overloading. ML fit is much better for any, even very small, ADC overloading and it does not require to pre-process the record and eliminate overloading samples. LS fit can give nearly the same results but requires excluding the overloaded samples from the record,
- even with pre-processed data used for LS fit, the error of the ML fit is somewhat smaller than that of the LS fit (Fig. 3).

The following experiments were focused on influence of the ADC noise that causes alternating of ADC output codes on estimation error. The results achieved for 8-bits ADC and 400 samples in record are shown in Fig. 4 and Fig. 5.

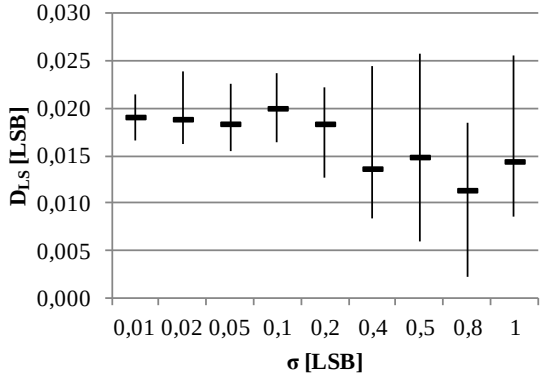


Fig. 4. Dependence of LS fit error on noise level.

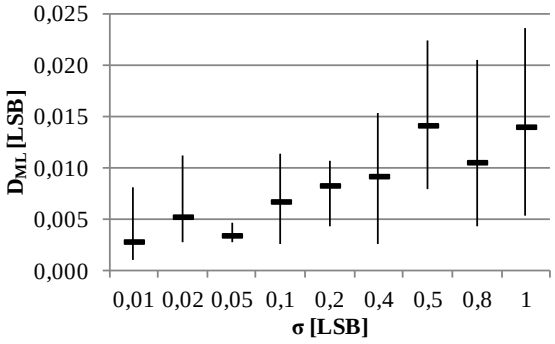


Fig. 5. Dependence of ML fit error on noise level.

ML fit is better for low noise and with increasing noise both fits give nearly the same estimation error. This can be explained by noticing that the input noise acts like dither, thus quantization noise is becoming more and more independent on the signal for larger nose amplitude. Uncertainties of results for repeated experiments are also comparable with mean values of the error as it was in previous cases.

The main goal of this comparative study was to qualify the effects of different fitting methods on precision of calculation of SINAD. To qualify the error, $SINAD_{REF}$ was calculated from precisely known stimulating sine wave and quantization noise achieved as a difference of record and precisely known stimulating sine wave by Eq. 4. Accordingly, $SINAD_{LS}$ and $SINAD_{ML}$ were calculated from sine wave recovered by LS and ML fit, respectively.

Fig. 6 and Fig. 7 show difference of $SINAD_{LS}$ and $SINAD_{ML}$ from $SINAD_{REF}$, respectively, as well as variance of repeated testing as a dependence on ADC noise σ. The results were achieved for linear 8-bit ADC and the record length $N=400$ samples taken by coherent sampling.

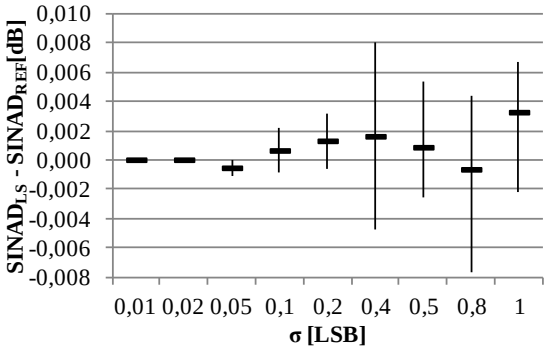


Fig. 6. Error of SINAD calculated from LS fit for different variance ADC noises.

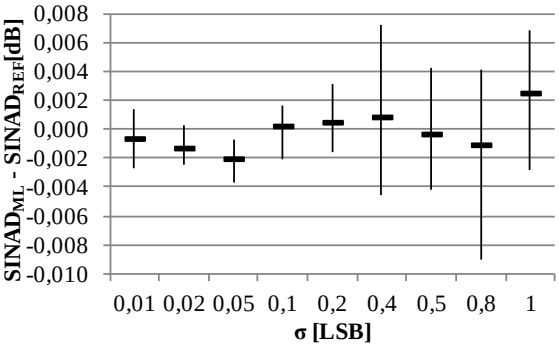


Fig. 7. Error of SINAD calculated from ML fit for different variance ADC noises. The SINAD error is larger for larger sigma partly because SINAD is measured with variance

The errors of SINAD calculated from both fits are very small and very similar. Increasing noise causes increasing of variance in results but this effect is also very similar for both methods.

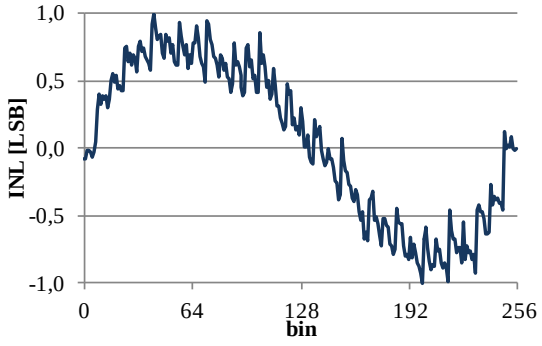


Fig. 8. Simplified INL of real ADC implemented in following simulation

Tests in the second group of evaluations were focused on effects on ADC nonlinearity at recovering the testing signals by LS and ML fits. To approximate as much as possible a real ADC in simulations, the INL measured on the real ADC in USB 6009 by National instruments was implemented in the simulation software. The real INL was simplified by rounding the lowest bits, in order to be equivalent to INL of a 8-bits ADC as it is shown in Fig 8. Approximate maximum likelihood estimates of the transition levels were determined from the sample record via the histogram method [1], applied

to the available 1024 samples, with the standard correction for the sine wave. In our view, this is the maximum information that can be extracted from the record for the *T(k)* values. During the fit, these estimates were kept constant, and the other 5 parameters were optimized.

Fig. 9 and 10 shows results of repetitive test achieved for different noise.

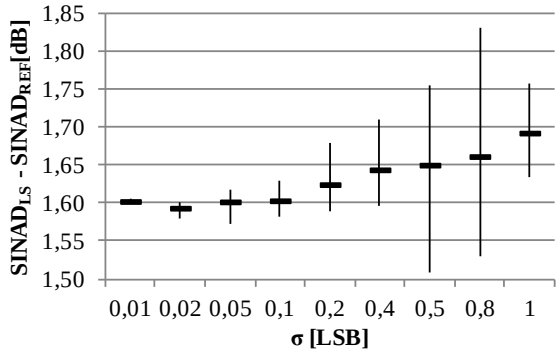


Fig. 9. Error of SINAD calculated from LS fit for different variance ADC noises.

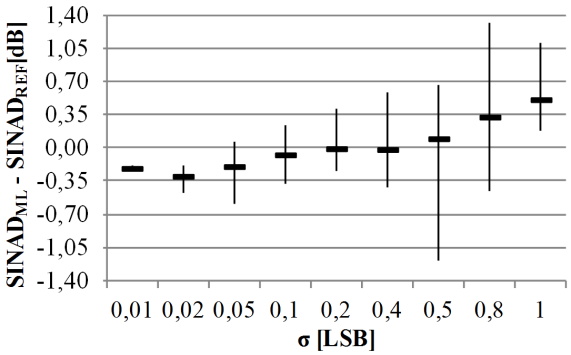


Fig. 10. Error of SINAD calculated from ML fit for different variance ADC noises.

ML fits is less sensitive on noise and the difference of calculated SINAD from reference one even decreases with increasing noise. On the other hand the variance of results from ML fit is higher than that from LS fit for high σ. Anyway the differences are small in order of 1 ÷ 2 dB.

IV. Conclusion

In this paper a comparative experimental study of LS and ML fits for ADC testing was carried out. Two main aspects were investigated: (i) effectiveness and precision of sinewave fit from record if ADC input range is overloaded and (ii) influence of fitting method on SINAD testing. The achieved results show that both methods give similar results for both studied conditions and applications. For overloading ADC ML fit may be directly applied for whole record while using LS fit the record must be first pre-processed by excluding the overloading samples. ML is not dramatically better, but it is better.

For SINAD testing where ADC input is not overloaded by the test signal, both fit methods give reasonable results. The differences are in the order of 1 ÷ 2 dB for LS, and 0.1 ÷ 0.3dB for ML.

Using the numerical minimization method based on the DE algorithm, the process of ML fit can be simply implemented in LabVIEW using built-in function `GlobalOptimization.vi`. For 1024 samples, the run time of the DE method was about 10 seconds on a standard PC, and a few seconds for the gradient-based method

Acknowledgement

This paper was written with support of DAAD agency in the frame of the project „IPID – International promovieren in Deutschland: Autonome Mikrosysteme für die Biosensorik, Bundesministerium für Bildung und Forschung".

The work is a part of the project supported by the Science Grant Agency of the Slovak Republic (No. 1/0555/11).

References

[1] IEEE Standard 1241-2011, "Standard for Terminology and Test Methods for Analog-to-Digital Converters", The Institute of Electrical and Electronics Engineers.

[2] IEEE Standard 1057-2007, "Standard for Digitizing Waveform Recorders", The Institute of Electrical and Electronics Engineers.

[3] D. Dallet and J. M. da Silva: "Dynamic Characterisation of Analogue-to-Digital converters." Springer 2005, ISBN-13 978-0-387-25902-4

[4] J. J. Blair: "Sine-fitting software for IEEE standards 1057 and 1241", in Proc. of the 16th IEEE Instrumentation and Measurement Technology Conference, IMTC/99, Venice, Italy, May 1999, vol. 3, pp. 1504-1506.

[5] I. Kollár and J. J. Blair: "Improved Determination of the Best Fitting Sine Wave in ADC Testing", IEEE Trans. on Instrumentation and Measurement, Vol. 54 No. 5, pp. 1978-1983, 2005.

[6] L. Balogh, B. Fodor, A. Sárhegyi, I. Kollár: "Maximum Likelihood Estimation of ADC Parameters from Sine Wave Test Data", In: Proc. of 15th IMEKO TC4 Symposium on Novelties in Electrical Measurements and Instrumentation: 12th Workshop on ADC Modelling and Testing. Iasi, Romania, 2007. (IMEKO), pp. 85-90.

[7] L. Balogh, I. Kollár, L. Michaeli, J. Šaliga, J. Lipták: „Full Information from Measured ADC Test Data using Maximum Likelihood Estimation". MEASUREMENT 45:(2) pp. 164-169. (2012)

[8] LabVIEW 2011 built-in help/guide

[9] Differential Evolution home page http://www.icsi.berkeley.edu/~storn/code.html

Contribution to the Study of Suitability of Different LP Filter Based on LTCC for I – Q demodulator

[1]*Kornel RUMAN (2st year),* [2]*Tibor ROVENSKÝ (1st year)*
Supervisor: [3]*Alena PIETRIKOVÁ*

[1,2,3]Dept. of Technologies in Electronics, FEI TU of Košice, Slovak Republic

[1]kornel.ruman@tuke.sk, [2]tibor.rovensky@tuke.sk, [3]alena.pietrikova@tuke.sk

***Abstract*— This paper deals with development, design and simulation of various LP (Low Pass) filters based on LTCC (Low Temperature Cofired Ceramics) for UWB (Ultra Wide-Band) sensor system. This paper demonstrates the design and full-wave electromagnetic simulation of two microstrip and one stripline LP filters using the HyperLynx 3D EM Designer (from Mentor Graphics). There are presented simulated results of insertion loss and return loss response of LP filters. It assesses the suitability of LTCC material system Green Tape 951 for the production of LP filters for high frequency area. The presented filters should be used as an antialiasing LP filter meant for I – Q demodulator that is presented in [1] and that would be one part of laboratory UWB sensor system.**

***Keywords*— LP, UWB sensor system, I – Q demodulator, LTCC, Green Tape 951.**

I. Introduction

Currently, the broadband technology plays very important role, in our lives. Very often, all around us, there are devices operating in the frequency range from 100 MHz to 10 GHz. In this frequency range operate various devices from mobile phones through television or satellite receivers and localization systems, sensor systems and so on.

UWB sensor systems can be used in a wide range of applications. We can find it in industry (detection of moisture or cracked pipes [2]), health (detection of cancer cells [3], detection of breathing [4]), agriculture (detection of larvae in the wood [5]), in military (underground mine detection, location objects, detection of people behind the wall [6]), emergency services (detection of people under the rubble) and many others.

With systematic research and development in the field of UWB sensor systems there are very often problems with the lack of relatively cheap, long-term stable and accurate LP filter. The big challenge is design, simulation and construction of a quality filter for removing unwanted frequencies. In particular, design and simulation of LP filters for UWB sensor system with cut-off frequency 3.5 GHz and a minimum attenuation -40 dB in stop band is the main challenging problem.

The progressive trends in the development of new materials for high frequency ranges tend to use LTCC ceramic for UWB sensor systems, thanks to its excellent mechanical and dielectric capabilities. Recent advance in new materials and process of production in the field of LTCC, caused rapid development of new filters, which may be part of more complex systems like I - Q demodulator in UWB radar sensor systems. Implementation of the required filters using microstrips or striplines and by using of stable dielectric properties of LTCC ceramics, ensure the stable behavior of electronic systems and increased integration of the system by creating 3D structure.

Advances in CAD (Computer-Aided Design), such as full-wave EM (Electromagnetic) simulators did coup in design of filters. In this paper is use for design of filters software HyperLynx 3D EM Designer, which facilitated the mentioned filters based on LTCC.

II. Filter design

A. Substrate

In the HF (High Frequency) area, the dielectric properties of substrate have a major impact on the quality and stability of the filter. Each substrate at the market has various parameters. This is the reason why not every substrate is proper for UWB area and choosing the appropriate substrates should pay close attention. Those with lower dielectric constants are more suitable for lower frequency applications. The FR4 (Flam Resistant based on epoxy and glass – FR4) laminate is the most common electronic carrier for circuits operating below the microwave frequency. Beyond 1 GHz, the laminate properties (like ε_r, tan δ, roughness, etc…) of the FR4 become nonlinear. It becomes difficult to precisely characterize the FR4 at the desired microwave frequency. Beyond 3 GHz, the use of the FR4 is not recommended anymore, in the case of unacceptable attenuation. Significant parameter in HF area is the value of dissipation factor Tan (δ) [7].

For correct design and simulation is necessary to know the parameters of the substrate. As substrate we chose LTCC material system Green Tape 951. Green Tape 951 is a low/temperature cofired ceramic tape. The 951 system comprises a complete cofireable family of Au and Ag metallization, buried passives, and encapsulants. 951 is available in multiple thicknesses and is designed for use as an insulating layer in multichip modules, single chip packages,

ceramic printed wiring boards and RF modules [8]. Parameters of used substrate are in Table I.

TABLE I
GREEN TAPE 951 PX SUBSTRATE SPECIFICATION [8]

Property	Units	Typical Value
Unfired Thickness	µm	254 ± 3
X, Y Shrinkage	%	12.7 ± 0.3
Z Shrinkage	%	15 ± 0.5
Surface Roughness	µm	< 0.34
TCE (25 to 300 ^{0}C)	ppm/^{0}C	5.8
Density	g/cm^3	3.1
Camber	(um/25 mm)	25
Thermal conductivity	(W/m.K)	3.3
Flexural strength	(MPa)	320
Dielectric constant (3 GHz)	-	7.8
Loss Tangent (3 GHz)	-	0.006

Low dielectric loss as typical property of LTCC material system Green Tape 951 allows using them in many applications where at high operating frequency are conventional laminates circuit boards limited.

B. Metallization

For correct design and simulation is necessary to know also the type and parameters of used metallization. For simulations of two microstrip filters we use parameters of conductor paste DuPont 6146. DuPont 6146 is an external solderable cofireable silver/palladium conductor compatible with Green Tape 951 low temperature cofired ceramic system [9]. For simulations of stripline filter we use parameters of solder paste DuPont 6145. DuPont 6145 is an internal solderable cofireable silver conductor compatible with Green Tape 951 low temperature cofired ceramic system [10]. Parameters of used solder pastes are in Table II.

TABLE II
SPECIFICATION

Property	6146	6145
Viscosity (Pa.S)	130 - 230	120 - 200
Dried Line Resolution (µm) line/space	125 / 125	125 / 125
Fired Thickness (µm)	8 -12	18 - 25
Fired Resistivity (mΩ/sq)	< 60	< 3

Both systems Dupont 6146 and 6145 are ideally suited for applications requiring excellent conductivity.

C. Initial Design

These filters were designed with cut-off frequency 3.5 GHz and a minimum attenuation -40 dB in stop band. We decided consider suitability of two microstrip filters and one stripline filter.

The first LP filter is realized as microstrip Stepped-Impedance (Figure 1a). The design parameters of *High-Impedance Line Width* (0.1 mm) and *Low Impedance Line Width* (4.5 mm) were optimized to meet the specifications in the pass band. Based on the results of the simulations, the optimum minimum number of 9 resonators was determined, in which filters meet the required min. attenuation of -40 dB at frequencies from 3.5 GHz to 10 GHz. The area allocated by the microstrip stepped impedance filter is approximately 35.61 by 4.5 mm.

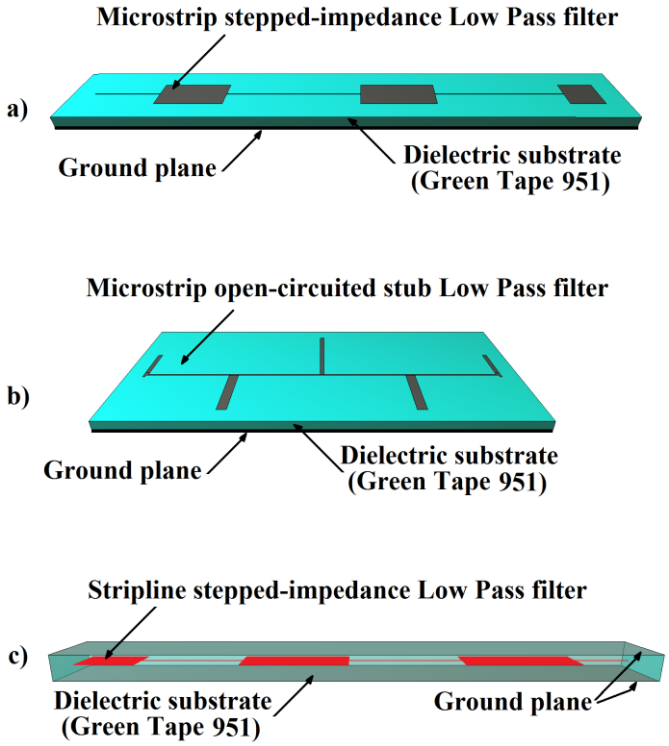


Fig. 1. 3D layout of three different LP filter: a) Microstrip Stepped-Impedance LP Filter b) Microstrip Open-Circuited Stub LP Filter c) Stripline Stepped-Impedance LP Filter

The second LP filter is realized as microstrip Open-Circuited Stub (Figure 1b). Based on the results of the simulations, the optimum minimum number of 9 resonators and geometric dimensions of microstrips were determined, in which filters meet the required min. attenuation of -40 dB at frequencies from 3.5 GHz to 10 GHz. The area allocated by the microstrip Open-Circuited Stub filter is approximately 33.85 by 11.53 mm.

The third low LP is realized as strip-line Stepped-Impedance (Figure 1c). The design parameters of *High-Impedance Line Width* (0.1 mm) and *Low Impedance Line Width* (5 mm) were optimized to meet the specifications in the pass band. Based on the results of the simulations, the optimum minimum number of 9 resonators was determined, in which filters meet the required min. attenuation of -40 dB at frequencies from 3.5 GHz to 10 GHz. The area allocated by the strip-line stepped impedance filter is approximately 30.74 by 5 mm.

The filters input and output are matched to 50 ohm characteristic impedance. The 50 ohms choice is a compromise between power handling capability and signal loss per unit length, for air dielectric [11].

III. RESULTS AND ANALYSES

The full-wave electromagnetic simulation of insertion loss (S_{21}) and return loss (S_{11}) of LP filters were purposely made up to 10 GHz to verify the correct function of the filters and suitability of the ceramic material Green Tape 951 at high frequencies.

A. *Insertion loss*

Simulated results of insertion loss of three different LP filters are shown in the Figure 2. The Figure 2 shows that the insertion loss of all tree designed LP filters meets the required min. attenuation of -40 dB at frequencies from 3.5 GHz to 10 GHz. Insertion losses of all designed LP filters are acceptable for use in the UWB applications. It is possible to use these filters as antialiasing LP filters in UWB devices such as IQ demodulator presented in [1] or part of system for Through Wall Moving Target Tracking by M-sequence UWB Radar presented in [12].

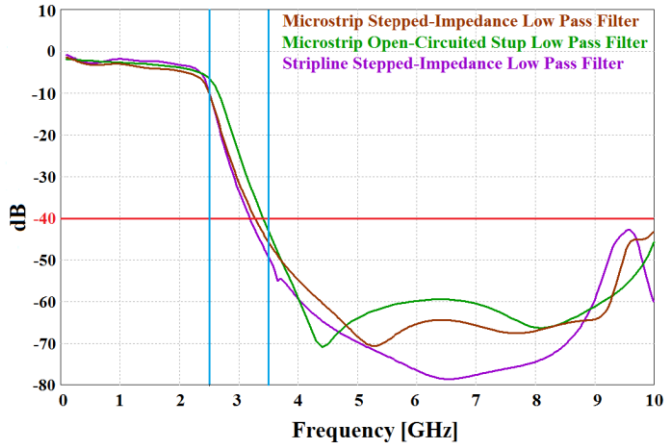


Fig. 2. Comparison simulated results of insertion loss of three different LP filters

B. *Return loss*

Simulated results of return loss of three different LP filters are shown in the Figure 3.

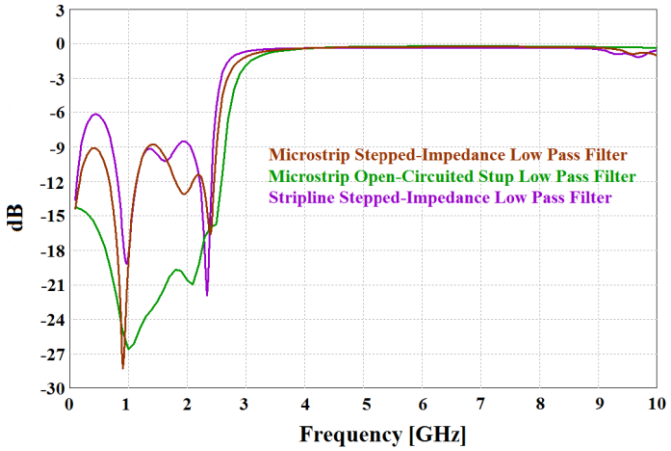


Fig. 3. Comparison simulated results of return loss of three different LP filters

Return los has sufficiently met the design criteria. Acceptable profile of insertion loss results in worse return loss but it was counted with this situation in simulation.

IV. Conclusion

The three LP filters with stop band frequency 3.5 GHz and a minimum attenuation -40 dB in stop band were designed and simulated. The layouts of LP filters were created by the HyperLynx 3D EM Designer (from Mentor Graphics) and simulated with full-wave electromagnetic simulation. Comparison between two microstrip and one stripline LP filters simulated results of insertion loss and return loss were presented. In these simulations, we verified the suitability of the ceramic material Green Tape 951 for the production of these types LP filters for high frequency area. Transmission characteristics of LP filters are acceptable for use in the UWB sensor system as an antialiasing LP filters for I – Q demodulator. This project was basis for manufacturing such filter based on LTCC multi-layer technology, and in the future work we will develop filters fabricated on Green Tape 951 and measure by Signal Analyzer N9020A.

Acknowledgment

This work was supported by VEGA Grant No. 1/0218/13.

References

[1] Žiga M., Liptaj M., "Design of the I-Q Demodulator for UWB Applications", SCYR 2012 - 12th Scientific Conference of Young Researchers - FEI TU of Košice, 2012, Košice.

[2] J. Sachs, A. Badstubner, F. Bonitz, M. Eidner, M. Helbig, R. Herrmann, M. Kmec, P. Rauschenbach, and H. Solas, "High resolution non-destructive testing in civil engineering by ultrawideband pseudo-noise approaches," in Ultra-Wideband, 2008. ICUWB 2008. IEEE International Conference on, vol. 2, sept. 2008, pp. 137 –140.

[3] M. Helbig, I. Hilger, M. Kmec, G. Rimkus, and J. Sachs, "Experimental phantom trials for UWB breast cancer detection," in Microwave Conference (GeMiC), 2012 The 7th German, march 2012, pp. 1 –4.

[4] E. Zaikov, J. Sachs, M. Aftanas, and J. Rovnakova, "Detection of trapped people by UWB radar," Microwave Conference (GeMIC), 2008 German, pp. 1 –4, march 2008.

[5] J. Sachs, R. Herrmann, M. Kmec, M. Helbig, and K. Schilling, "Recent Advances and Applications of M-Sequence based Ultra-Wideband Sensors," in Ultra-Wideband, 2007. ICUWB 2007. IEEE International Conference on, sept. 2007, pp. 50 –55.

[6] D. Urdzik, D. Kocur, and J. Rovnáková, "Detection of multiple targets with enhancement of weak UWB radar signals for the purposes of through wall surveillance," in Applied Machine Intelligence and Informatics (SAMI), 2012 IEEE 10th International Symposium on, jan. 2012, pp. 137 –142.

[7] Nikholas G. Toledo, "Practical Techniques for Designing Microstrip Tapped Hairpin Resonator Filters on FR4 Laminates", in Communications Engineering Division, Advanced Science and Technology Institute, Diliman Quezon City, Philippines 1101

[8] Dupont Microcircuit Materials, Design and Layout Guidelines, Available on the Internet:,4.1.2013, http://www2.dupont.com/MCM/en_US/assets/downloads/prodinfo/GreenTape_Design_Layout_Guidelines.pdf

[9] Dupont Microcircuit Materials, Technical Data Sheet Dupont 6146, Available on the Internet:,4.1.2013, http://www2.dupont.com/MCM/en_US/assets/downloads/prodinfo/6146.pdf

[10] Dupont Microcircuit Materials, Technical Data Sheet Dupont 6145, Available on the Internet:,4.1.2013, http://www2.dupont.com/MCM/en_US/assets/downloads/prodinfo/6145.pdf

[11] MicrowavesLOL.com, Why 50 ohms?, Available on the Internet:, 4.1.2013, http://www.microwaves101.com/encyclopedia/why50ohms.cfm.

[12] Rovnakova J., Kocur D., "UWB Radar Signal Processing for Through Wall Tracking of Multiple Moving Targets", 7th European Radar Conference, 2010, 204-207 p.

Control of Servodrives for BIOLOID Robot with GUI interface

[1]Radovan SIVÝ (1st year), [2]Ján BAČÍK (1st year)
Supervisor: [3]prof. Ing. Daniela PERDUKOVÁ PhD.

[1,2,3]Department of Electrical Engineering and Mechatronics, FEI TU of Košice, Slovak Republic

[1]radovan.sivy@tuke.sk, [2]jan.bacik.2@tuke.sk, [3]daniela.perdukova@tuke.sk

***Abstract* — This paper describe Graphical User Interface for controlling a group of drives moving a robotic structure. Graphical User Interface (GUI) enables to control hexapod robot to create own movement sets and change parameters of servodrives. Paper also explains communication principles with the servodrives, development of digital packets needed for its control as well as development of control structure that is based on S-functions in the C-MEX language in the MATLAB environment.**

***Keywords* — BIOLOID, MATLAB, Simulink, control, GUI, S-functions, servodrive, Dynamixel**

I. Introduction

Nowdays robotics is not just common part of industrial production, but also becomes a part of everyday life. While industrial robots have a fixed location and their workspace, mobile robots moves on wheeled or tracked chassis. One of the possible solutions of movement is inspired by movement of various creatures in nature, such as animals or insects usingfor motion their set of legs. Musculoskeletal system of animals enables to overcome difficult obstacles placed on the ground. This is the main reason for using this approach in mobile robotics. Therefore, this paper is dedicated to a six-leged robot - hexapod. This type of chassis has advantage against wheel or tracked chassis such as high stability and dexterity. Drawback of this chassis type consist in the complexity of control due to regulation of each leg separately. Control of the hexapod is described in several papers [1],[2]. The main issue of our work was to develop a program using serial communication to control each servodrive so that the robot is able to perform the required actions and movements. For this solution we have chosen the program MATLAB, which contains necessary tools to create and communicate graphical interfaces [3].

II. Description of BIOLOID kit

BIOLOID is a platform developed by Robotis company (www.robotis.com) for educational purposes in robotics. It consists of modular construction components and small servodrives of the Dynamixel AX-12A type, which can be connected arbitrarily, according to requirements. This allows assembling different robot configurations, such as wheeled, legged, or humanoid types of robots. Hexapod-type structure (Fig. 1) is used in our work.

Fig. 1 Hexapod-type robot

The Dynamixel module consists of a DC motor, gear, series of sensors (to sense temperature inside the motor, motor torque, speed and position) and a board with digital control circuit. These elements enable to construct an intelligent drive module designed to fulfil complex tasks and to build an intelligent servodrive. Except for the communication, the control circuits monitor physical variables of the servodrive (temperature, torque, and voltage) and evaluate their state depending on limitations defined by a user. The safety measures enable the control system to switch off the servodrive at exceeding monitored variables over a limiting pre-set value(s). The servodrive itself can be controlled with a comparatively good precision because resolution of the position sensor presents 1024 steps per revolution which is enough for performing common tasks by a robot.

III. Develop of Graphical User Iinterface

The GUI MATLAB environment includes a basic graphical elements (buttons, lists, text fields, etc.), that enable to develop an user interface to control any system. To create the control panel, we need to run an environment for GUI in program MATLAB by command *guide* in the command line. Below, we describe some of the main parts of our GUI

a) Movement Maker

Movement Maker is utility in the first part of the GUI environment (Mmaker). It serves for creating own movements of robotic structure. Mmaker uses external file that contains rotations of servodrives in each step. Repeating these steps creates a motion cycle. Each step of motion cycle is created by accessing actual position of each servodrive, These data are saved to an external file. Mmaker contains five control buttons:

READ – reads the current position of servodrives. Where each value is in the table (Fig. 2).

WRITE – records the current location to an external file

CLEAR – delete a file stored positions

RUN – start cycle movements

STOP – stop cycle of movements

The tool also includes *speed* editing window that allows you to specify the speed of the movement cycle. Editing window *Repeat* indicates repeats of movement cycle. Entering a value of 0, the movements are repeated until the STOP button is pressed.

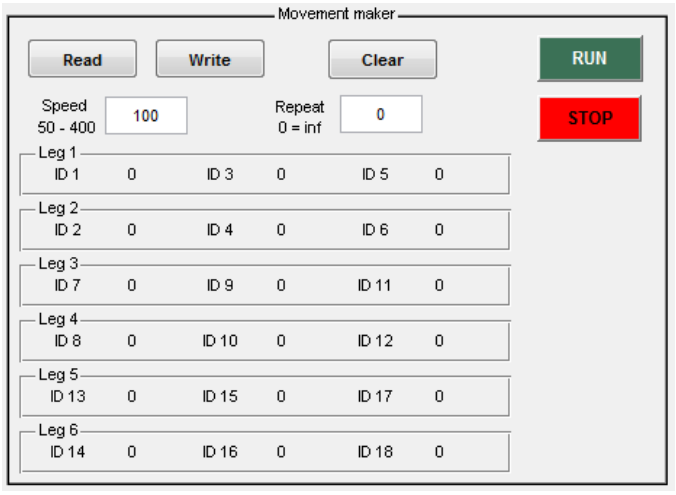


Fig. 2 The GUI subscreen of the Movement Maker

b) Control of hexapod movements

Next part of the GUI environment shown in Fig. presents a control panel for hexapod movement. It contains buttons that represent different movements of the robot. These moves are designed only for one robot type, in our case the hexapod. When pressing *Move* button, it starts m-file, which cycle of movement until the movement stops or a different type of the movement is chosen. This panel also contains a slider and edit box for change of movement speed. This change can also be performed during robot movement. Speed value range is from 5 to 45 rpm. Another slider is used to change the height of the hexapod. Change of height does not affect individual movements. The control panel also includes switch of pre-programmed motion modes.

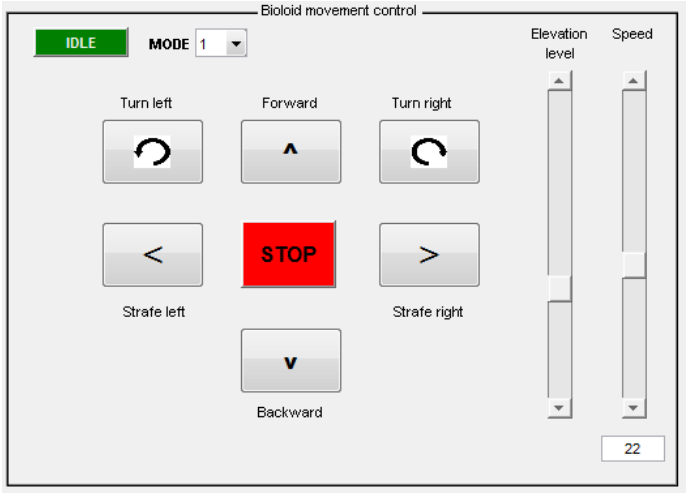


Fig. 3 The GUI subscreen of Bioloid control

c) Dynamixel control

The operational screen for setting parameters of a servodrive is shown in Fig. 4. After selecting a chosen servodrive its actual parameters are read and shown in the movable sliders and in editing boxes. Reading the parameters is actuated by requirement of the control programme written in the m-file (i.e. by instructions of the "file" type).

Three groups of sliders are here:

1) The sliders in the first group set up position and speed of a chosen drive.
2) The sliders in the second group change the lower and upper angle limit.
3) The sliders in the third pair set both the maximal torque developed by the servodrive and value of maximal temperature for switching off the servodrive (in order to protect the servodrive against overloading).

The instruction packets for the pairs of sliders were developed in Simulink model is using a S-function. Communication through the S-functions is described in next chapter.

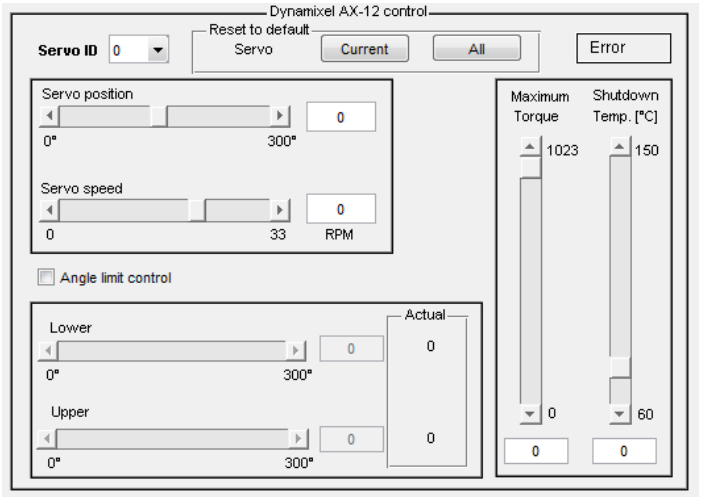


Fig. 4 The GUI subscreen of Dynamixel drives control

IV. COMMUNICATION WITH SERVODRIVES

Communication with servodrives from MATLAB environment is realised by the instructions of the "file" type (*fopen*, *fclose*, etc.) which concern the work with files (reading. writing, etc.). In special case they can be set to a mode that uses the serial port for communication. In our case, when working in the GUIDE (the GUI development environment), it is suitable to define a global variable *handles.s* as a *serial(obj)* and insert parameters for communication as shown here:

```
handles.s = serial(COM1);
set(handles.s,'BaudRate',
200000,'Parity','None','DataBits',8, 'StopBits', 1);
```

The variable with defined parameters of serial communication is inserted into the instruction *fopen*, that opens the specific serial port:

```
handles.s = serial(COM1);
```

The functions *fwrite* and *fread* enable sending and receiving the data through the serial port. Here under term "the data" we have in mind an instruction packet for the Dynamixel servodrive. Due to calculation of a checksum the packet originally developed in the 16-bit system has to be re-calculated to the 10-bit system. Using the instruction *fwrite* the packet starts to travel along the communication line. On other side the instruction *fread* is used to read the state packet. When handling with more servodrives at once it is necessary to develop more similar functions.

After finishing the communication it is recommended to close the opened communication channel what is done by instruction *fclose(handles.s).*

When starting the GUI more times a duplication of the serial channels in the programme MATLAB can occur. To avoid this fault, the instruction *fclose(instrfind)* is used. Which finds all previously opened serial connections that have remained in the memory and closes them.

The basic idea of this communication type consists in development of the control programme using an S-function written in the C-MEX language and using instructions from the MATLAB Instrument Control Toolbox for communication with external serial equipment. In the Simulink model the S-function presents a block containing various equations and calculations of parameters, [5]. In our case the S-function ensures development of an instruction packet for control of the Dynamixel servodrives. In this case it is necessary to choose properly the blocks from the Instrument Control Toolbox and to set up them in such a way so that the serial communication would work perfectly.

The S-function presents a dynamic block whose the outputs are function of sampling period, inputs, and states. The S-functions use a syntax enabling to communicate with Simulink during the operation. Their form is enough simple and they are suitable both for discrete, continuous and hybrid systems.

In principle, more-or-less each model in Simulink can be described by the S-function. The S-functions can be written directly in the MATLAB (m-file) or as a C-MEX functions using a programming language (the best in the C/C++ language but also other ones can be used, e.g. Ada or Fortran). What concerns a comparison, the S-functions of the C-MEX type have the same structure but here they offer better possibilities and more possibilities and more functions like the functions of the M-file type.

The C-MEX programme defined in an S-function block must contain information about model structure during the whole simulation period. The S-functions written in the C-language is compiled as Mexfiles using the instruction *MEX*. In our case for the compilation we have used the compiler OpenWatcom 1.9.

For development of the S-functions it is necessary to preserve a sequence of the particular states (*Initialisation, Calculation of the following sample, Calculation of outputs, Refreshing of disc states, Calculation of derivations, Finishing simulation task*).

V. CONCLUSION

The developed GUI and methods of servodrives control were applied for control of movement of a robotic hexapod structure [4]. The procedure can be extended for driving of any robotic structure type developed by the Robotis company. Communication between the GUI environment and the hexapod is ensured trough USB2Dynamixel that provides serial data transfer via USB port on the computer. The link was realized using TTL logic directly connected to the communication line of servodrives. For motion control of the hexapod-type robot we used serial interface develop in MATLAB by "file" type instructions. Individual servodrives of the robot were controlled by the Simulink model with C-Mex S-function.

ACKNOWLEDGMENT

This work was supported by the Slovak Research and Development Agency under the contract No. APVV-0185-10.

REFERENCES

[1] TEODORO, Pedro: Development of a simulation environment of an entertainment humanoid robot. Dissertation thesis, TU Lisbon.

[2] VARTAŠ, Tomáš.: Control Motion of Humanoid Type Robot. Diploma thesis, FEEaI TU Košice, 2010 (in Slovak).

[3] Developing S-Functions. Available at: http://www.mathworks.com/help/pdf_doc/simulink/sfunctions.pdf

[4] SIVÝ, Radovan: Control of Servodrives for the Robot BIOLOID, Diploma Thesis, FEEaI TU Košice, 2012 (in Slovak).

[5] Developing S-Functions. Available at: <http://www.mathworks.com/help/pdf_doc/simulink/sfunctions.pdf>

Creation of a 3D Robot Manipulator Model for Use in MATLAB

[1]*Marek VACEK (2st year – PhD.),* [2]*Martin LEŠO (1st year – Ing.)*
Supervisor: [3]*Jaroslava ŽILKOVÁ*

[1]Dept. of Electrical Engineering and Mechatronics, FEI TU of Košice, Slovak Republic

[1]marek.vacek@tuke.sk, [2]martin.leso@student.tuke.sk, [3]jaroslava.zilkova@tuke.sk

***Abstract*—The following article deals with the creation of a visual robotic arm model which can be used in MATLAB. The article describes the creation of the 3D model, its exporting, further processing and the final phase of the import, modification of the model and its use in MATLAB itself.**

***Keywords*—robot manipulator, robotic arm, 3D model, Robotics Toolbox, Graphical user interface**

I. Introduction

In the academic and research environment, there is a constant demand for simulations of the already existing or newly created systems. In robotics, there are lots of programs which give us numerous options for simulation, whether visual or mathematical. However, the problem usually is that few programs provide us with suitable simulation options for both areas. There are several programs which can simulate visually and mathematically, but their proficiency is not the same in both areas. If the user needs mathematical simulation, MATLAB is the best choice.

MATLAB is a program for scientific numerical calculations, modeling, algorithm design, computer simulations, and analysis and presentation of data. As to visualization of 3D objects, it has toolboxes suitable for simple modeling, but modeling of more complex objects is very complicated and lengthy. It is, therefore, better to use other programs to create and modify them. However, MATLAB can work with them after they have been correctly created and exported.

MATLAB includes the patch function to visualize 3D objects. This function enables the displaying of 3D models of objects based on the vertices and faces entered. As a visual model of a robot manipulator consists of a vast number of vertices and faces (hundreds, even thousands), entering the values into matrices manually would be too complicated and lengthy. Therefore, the robot was firstly created in ProEngineer, then modified in Blender, and finally exported into and processed in MATLAB.

II. Creation of the 3D robot model in a CAD-type software

Firstly, a kinematic model of the robot has to be created in MATLAB using the Robotics Toolbox. This model will be referred to as RT_robot. Next, a graph showing all individual parts of the robot at zero position of joints can be displayed using the plot method (Fig.1).

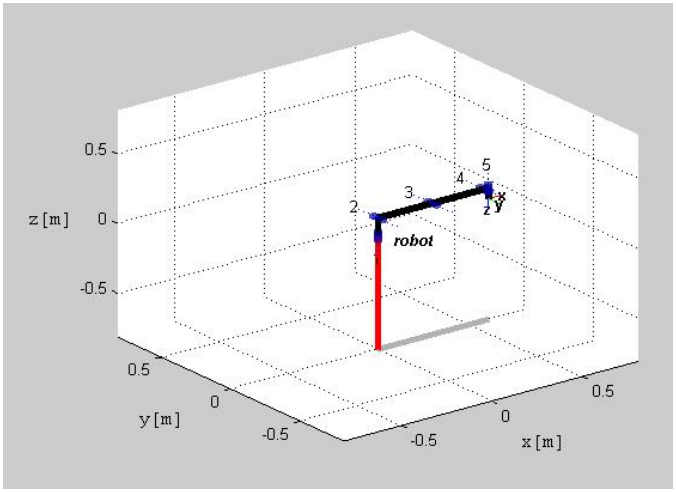


Fig. 1. Kinematic model of the robotic manipulator (MATLAB)

To simplify the design and modeling of the robotic manipulator it is useful to use CAD-type (Computer Aided Design) programs. There are several options from various developers – AutoCad, Pro Engineer, or Solid Works.

We created the robotic manipulator model in ProEngineer. The creation consisted of the design and modeling of all individual parts which were joined into one unit (robot configuration) in the final phase (Fig.2). Due to limited extent of this article, it is not possible to describe the modeling process.

To enable further processing, the 3D robotic manipulator model had to be exported into the .obj format.

III. Modification of the 3D robotic manipulator

As the exported file contains not only information about the robot model but information unsuitable for MATLAB as well, and as ProEngineer does not include tools necessary for modification of .obj files, this file (the 3D model) had to be modified in another program for modifying 3D models – Blender.

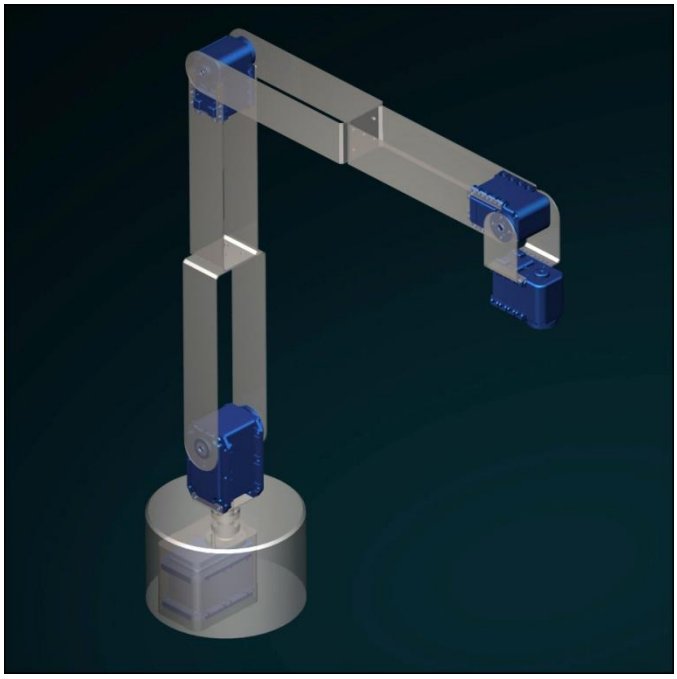


Fig. 2. 3D model of the robotic manipulator (ProEngineer)

The .obj model was imported into the program using the Import file function to finalize the modification of the visual model. In Blender, the robot had to be divided into separate parts (objects) which were numbered from 1 to n+1. Part 1 is the robot's base. This base is stationary. Part n+1 is the robot's effector.

In Blender, objects are selected using the right mouse button. If several objects need to be selected at the same time, the shift button has to be pushed and held while selecting the desired objects using the right mouse button. The ctrl+j shortcut will join the objects. The following picture shows the arrangement of the individual parts of the robot (different parts have different colors) (Fig.3):

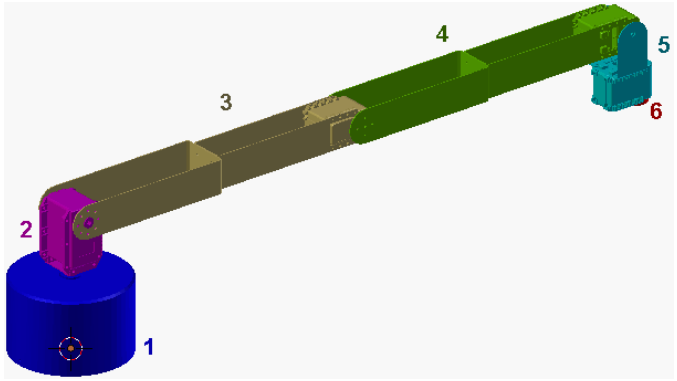


Fig. 3. 3D model of the robotic manipulator (Blender)

Besides naming and creating the individual parts of the robot it is also useful to modify the robot in such a way that the parts do not contain any redundant vertices and faces, adjacent faces are joined, and the object has suitable topology. Modified in this way, the model is exported into an .obj file again.

IV. Processing of the Model in MATLAB

The readrobot(path) function was used to import the .obj file into MATLAB. This function loaded the data saved in the .obj file and wrote them into the given structure. In our case, the robot structure had been saved under the name "robot". It was necessary to enter an absolute or a relative path to the file into the argument "path". It was then possible to display the robot in individual parts in MATLAB in the following way:

```
patch(robot.Link{i}.l)
```

where *i* is a numeric variable representing that part of the robot which the user wants to display. If the user wants to display the whole robot, the following commands must be entered:

```
figure(1)
for i=1:6
patch(robot.Link{i}.l)
hold on
end
```

The resulting robot was displayed without shading and in a distorted scale. That is why the following commands were used to improve the displaying of the robot:

```
grid on          % adding the grid
axis equal       % equal axes
camlight         % adding lighting
```

After entering the commands, the robot was displayed as follows (Fig.4):

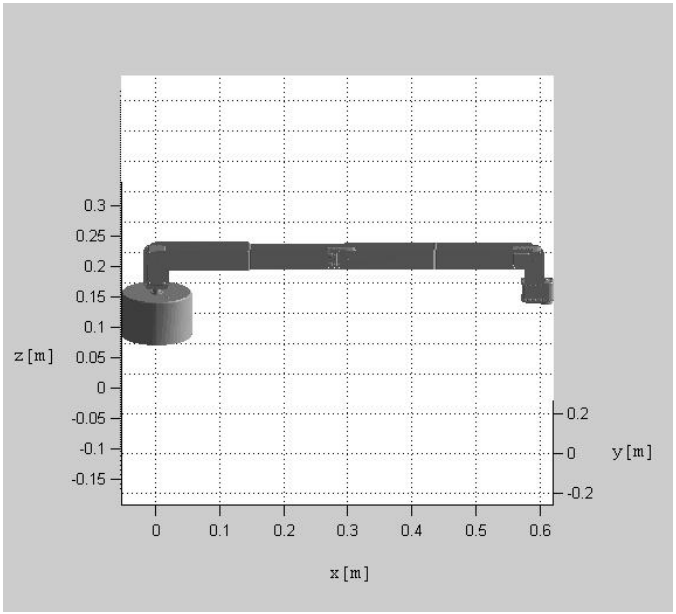


Fig. 4. 3D model of the robotic manipulator (Matlab)

If the user wants to change the robot's position by entering values for particular joints of the robotic manipulator (forward kinematics calculation), the vertices of the model have to be

recalculated. That is why auxiliary vertices were created, which, when multiplied by the transformation matrix of the robot link, give us the required position of the robotic manipulator link.

Transformation matrices of individual links for the given values of the joint variables were determined as follows:

```
  T{1}=RT_robot.base;
  n = RT_robot.n;
  for i=1:n
  T{end+1}=T{end}*RT_robot.links(i).A(q(i)
);
  end
```

As the displayed robot's joint variables had values set at zero, the row vector was defined as follows:

```
  q=[0 0 0 0 0];
```

The dimensions of the resulting transformation matrices are 4x4. The vertices of the robot links are in the shape of m*3, where m is the number of vertices in a link. Therefore, a column vector was added by an auxiliary vertex, and the whole matrix was transposed and then multiplied by an inverse transformation matrix T. In MATLAB, this was done as follows:

```
  for i=1:n+1
  vertices=[robot.Link{i}.l.vertices
ones(length(robot.Link{i}.l.vertices),1)]
.'
  robot.Link{i}.vertices=(inv(T{i})*vertic
es);
  end
```

After creating the auxiliary vertices of the robot links – robot.Link{i}.vertices – any position of the robot could be calculated in the following way:

```
  q=[ 0 pi/2 -pi/2 0 0];
  clear T
  T{1}=RT_robot.base;
  n = RT_robot.n;
  for i=1:n
  T{end+1}=T{end}*RT_robot.links(i).A(q(i)
);
  end
  for i=1:6
  robot.Link{i}.l.vertices=(T{i}*robot.Lin
k{i}.vertices).';
  robot.Link{i}.l.vertices(:,4)=[ ]
  end
```

It was then possible to display the final position of the robot as follows:

```
  grid on          % adding the grid
  axis equal       % equal axes
  camlight         % adding lighting
```

The following picture shows the calculated position (Fig.5):

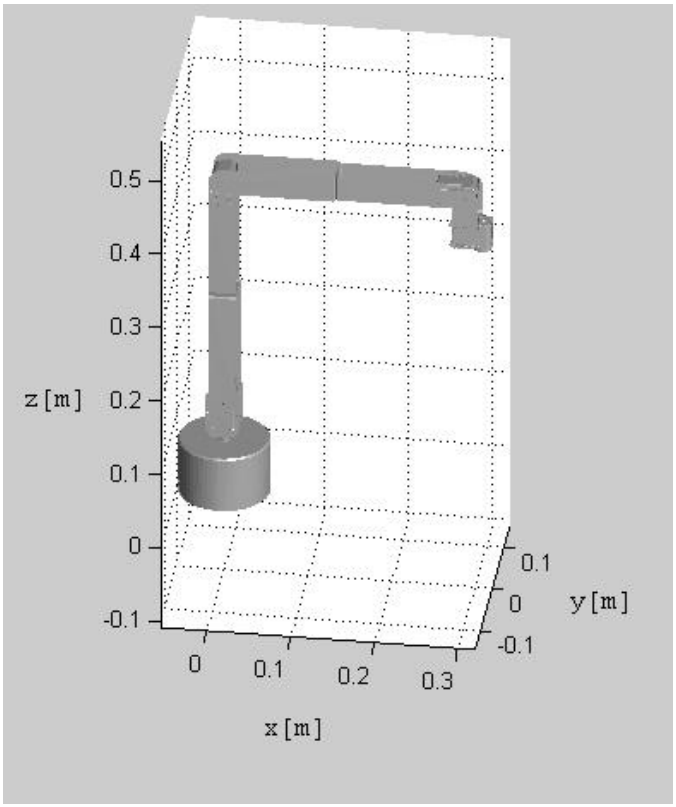


Fig. 5. Final visualization of the 3D model of the robotic manipulator (Matlab)

V. Conclusion

This article described the creation, modification, and use of a 3D robotic arm model suitable for further use in MATLAB. Three tasks had to be performed to create the model: create a 3D model of a robotic arm in a CAD-type program, export, modify, and simplify the .obj file, and import the file into and process it in MATLAB.

Acknowledgment

The work has been supported by project KEGA 011TUKE-4/2013.

References

[1] VACEK,M.: Modelovanie priemyselných robotov, Technical university of Košice, Faculty of Electrical Engineering and Informatics, Diploma Thesis, 2011
[2] VACEK,M.: Optický bezpečnostný systém robotického pracoviska, Technical university of Košice, Faculty of Electrical Engineering and Informatics, Postgraduate Thesis, 2013
[3] LEŠO,M.: Vizualizácia robotov, Technical university of Košice, Faculty of Electrical Engineering and Informatics, Bachelor Thesis, 2012
[4] Mathworks MATLAB: R2012a Documentation [online]. Available on the Internet: <http://www.mathworks.com/help/techdoc/>.
[5] Corke, Peter: Robotics Toolbox for MATLAB, Release 9. Available on the Internet: <http://www.petercorke.com/RTB/robot.pdf>.
[6] Corke, Peter: Robotics, Vision and Control. Springer-Verlag Berlin Heidelberg 2011. 570 s. ISBN 978-3-642-20143-1.
[7] Reza N. Jazar: Theory of Applied Robotics. Springer New York Dordrecht Heidelberg London 2007. 893 s. ISBN 978-1-4419-1749-2.

Determination of inductance levels of switched reluctance motor

[1]*Nataliia ISTOMINA (5th year),* [2]*Martin BAČKO (4th year)*
Supervisor: [3]*Andrii KALINOV*

[1,3]Institute of Electromechanics, Energy Saving and Control Systems,Kremenchuk Mykhailo Ostrohradskyi National University
[2]Dept. of Theoretical Electrical Engineering and Electrical Measurement, FEI TU of Košice, Slovak Republic

[1]istominaNM@yandex.ua,[2]martin.backo@tuke.sk

***Abstract*— The analytic dependence for calculation of maximum and minimum inductance levels under constant current using rated data of switched reluctance motor is obtained.**

***Keywords*— switched reluctance motor, variable phase inductance, inductance level.**

I. Introduction

At present the application of switched reluctance motor (SRM) in all technological fields is extend [1]. According to [2] switched reluctance drive (SRD) is prospective electric drive.

Operating mode of one SRM phase describes as shown on fig. 1.This mode can be afford by next stages:

Point 1: Supply voltage U feed per work phase.

Segment 1-2: Transient process of current I leap up. Stator tooth attracts rotor tooth by increased flux linkage Ψ.

Point 2: Stator and rotor tooth start to overlap. Phase inductance L start to increasing. Counter electromotive force (EMF) is induced at phase winding.

Segment 2-3: Current slowly decreases because EMF influence.

Point 3: Supply voltage disconnected or reverse voltage applied.

Segment 3-4: Phase inductance reaches maximum level. Current and flux linkage are decrease.

Point 3: Phase operating mode per commutation cycle is completed.

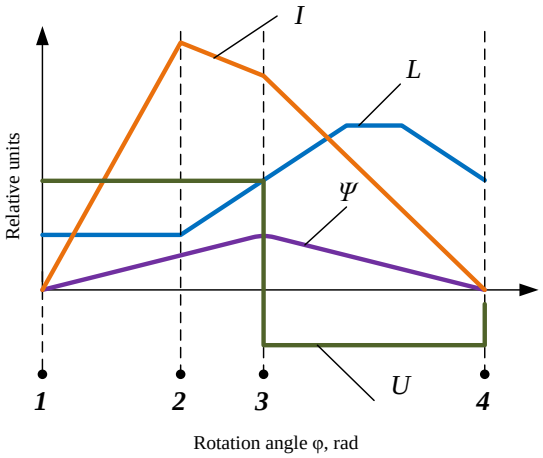


Fig. 1. Illustration of operating mode for one phase.

So commutation of phase operating modes depends on phase inductance. As shown in works [3–5] phase inductance is described by nonlinear dependence on rotation angle (Fig. 2). It is accepted to approximate such composite dependence by linear spline function, as shown on fig. 2.

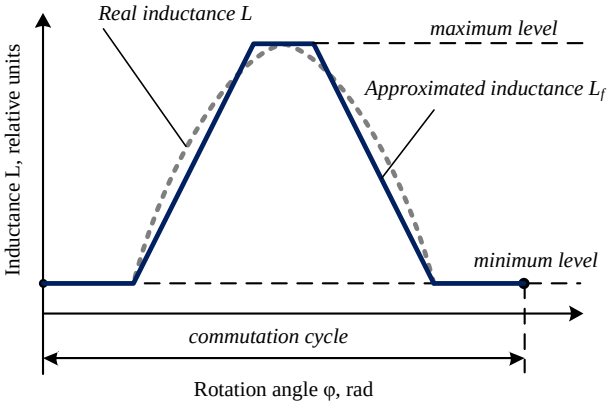


Fig. 2. Approximation of phase inductance dependence on rotation angle.

In works [6, 7] maximum and minimum levels of inductance is defined by experiments. It takes many working time.

The actual problem is determination of inductance levels using rated data of SRM.

II. Objective

Decreasing effort of mathematical modeling of switched reluctance motor by inductance mathematical definition.

III. Materials and Findings

Form of approximated inductance depends on (Fig. 3):

- inductance levels – L_{max}, L_{min};
- design factors: stator pole pitch β_S, rotor pole pitch β_R, pole pitches difference $\Delta\beta$,

$$\Delta\beta = \beta_R - \beta_S . \quad (1)$$

If $\beta_S = \beta_R$, then zone of inductance maximum level degenerates to point.

Slope of linear approximation,

$$a = \frac{L_{max} - L_{min}}{\beta_R} . \quad (2)$$

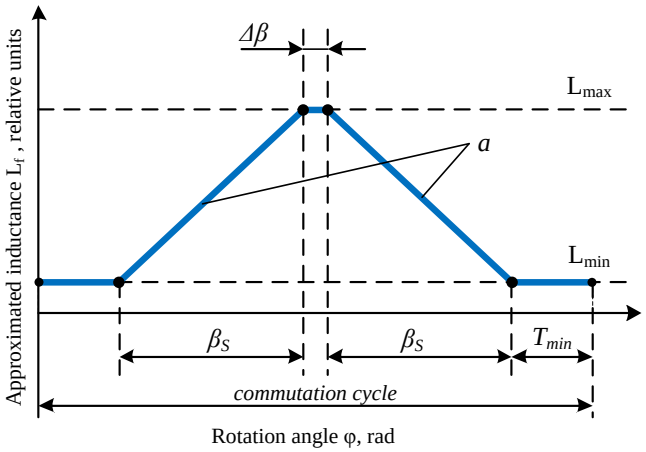


Fig. 3. Inductance dependence on rotor angle.

According to [8] maximum flux linkage SRM at unsaturated systems is

$$\Psi_{max} = \frac{\pi U_n}{Z_R \omega_n} , \quad (3)$$

where U_n – rated supply voltage,

$$U_n = \frac{P_{2n}}{\eta_n I_n} ;$$

P_{2n} – rated power output;

η_n – rated efficiency;

I_n – rated current;

Z_R – number of rotor poles;

ω_n – rated angular frequency,

$$\omega_n = \frac{2\pi n_n}{60} ;$$

n_n – rated angular velocity.

According to [3] inductance of consistent position notably maximum inductance level is,

$$L_{max} = \frac{\Psi_{max}}{i} . \quad (4)$$

where Ψ_{max} – maximum flux linkage;

i – instantaneous value of current,

$$i = I_m \sin \varphi_{el} = \sqrt{2} I_n \sin \varphi_{el} ;$$

φ_{el} – electrical rotation angle,

$$\varphi_{el} = \frac{\pi \omega t}{\varphi_k} ;$$

ω – instantaneous angular frequency;

φ_k – commutation angle,

$$\varphi_k = \frac{2\pi}{m Z_R} ;$$

m – phase number of motor.

Under constant current maximum level of inductance is,

$$L_{max} = \frac{\Psi_{max}}{\sqrt{2} I_n} . \quad (5)$$

As shown in [9], energy of magnetic field

$$W = \frac{\Delta L I_m^2}{2} = \Delta L I_n^2 . \quad (6)$$

On the other part, according to [10], energy of magnetic field, what expends on motor rotation for an angle $\Delta\varphi$ is,

$$W_R = M_n \Delta\varphi . \quad (7)$$

So we can take

$$M_n \Delta\varphi = \Delta L I_n^2 . \quad (8)$$

Under operating mode of one phase rotation angle is $\Delta\varphi = \varphi_k$, then inductance level difference

$$\Delta L = \frac{M_n \varphi_k}{I_n^2} . \quad (9)$$

So inductance of inconsistent position notably minimum inductance level is

$$L_{min} = L_{max} - \Delta L . \quad (10)$$

We obtain formulas for inductance levels determination:

$$\begin{cases} L_{max} = \dfrac{1}{\sqrt{2}} \dfrac{M_n}{I_n^2} \dfrac{\pi}{Z_R}, \\ \Delta L = \dfrac{2}{m} \dfrac{M_n}{I_n^2} \dfrac{\pi}{Z_R}, \\ L_{min} = \left(\dfrac{1}{\sqrt{2}} - \dfrac{2}{m} \right) \dfrac{M_n}{I_n^2} \dfrac{\pi}{Z_R}. \end{cases} \quad (11)$$

We obtain inductance levels by declared method for SRM with the following rated data:

- number of stator tooth, $Z_S = 6$;
- number of rotor tooth, $Z_R = 4$;
- power output, $P_{2n} = 3$ kW;
- phase current, 7.5 A;
- angular velocity, 3000 revolutions per minute;
- efficiency, 88 %;
- inertia moment, 0.0019 kg·m^2;
- electromagnetic torque, 9.5 N·m.

Calculations of design and commutation factors:

– commutation angle,

$$\varphi_k = \frac{2\pi}{m Z_R} = \frac{2 \cdot 3.14}{3 \cdot 4} = 0.524 \text{ rad, or } 30 \text{ deg};$$

– commutation period,

$$T = \frac{2\pi}{Z_R} = \frac{2 \cdot 3.14}{4} = 1.571 \text{ rad, or } 90 \text{ deg};$$

– absolute torque zone,

$$\gamma = \frac{\pi}{Z_R} = \frac{3.14}{4} = 0.785 \text{ rad, or 45 deg;}$$

–relative overlap zone,

$$\rho_a = \frac{\gamma_a}{\varphi_k} = \frac{45}{30} = 1.5;$$

–relative efficient overlap zone,

$$\rho_E = \frac{m}{2} \cdot \frac{Z_S}{Z_R} = \frac{3}{2} \cdot \frac{6}{4} = 2.25;$$

–stator pole pitch,

$$\beta_S = \frac{1}{2} \cdot \frac{2\pi}{Z_S} = \frac{1}{2} \cdot \frac{2 \cdot 3.14}{6} = 0.524 \text{ rad, or 30 deg;}$$

–rotor pole pitch is selected of design data tables, we take $\beta_R = 34$ deg,

–pole pitches difference,

$$\Delta\beta = \beta_R - \beta_S = 34 - 30 = 4 \text{ deg;}$$

–period of minimum inductance level,

$$T_{min} = T - 2\beta_S - \Delta\beta = 90 - 2 \cdot 30 - 4 = 26 \text{ deg.}$$

Calculations of rated data:

–rated supply voltage,

$$U_n = \frac{P_{2n}}{\eta_n I_n} = \frac{3000}{0.88 \cdot 7.5} = 454.545 \text{ V;}$$

–rated angular frequency,

$$\omega_n = \frac{2\pi n_n}{60} = \frac{2 \cdot 3.14 \cdot 3000}{60} = 314.159 \text{ rad/sec.}$$

–maximum flux linkage,

$$\Psi_{max} = \frac{\pi U_n}{Z_R \omega_n} = \frac{3.14 \cdot 454.545}{4 \cdot 314.159} = 1.136 \text{ Wb.}$$

Calculation of inductance levels:

–inductance of consistent position,

$$L_{max} = \frac{\Psi_{max}}{\sqrt{2} I_n} = \frac{1.136}{\sqrt{2} \cdot 7.5} = 107.137 \cdot 10^{-3} \text{ H;}$$

–inductance level difference,

$$\Delta L = \frac{M_n \varphi_k}{I_n^2} = \frac{9.5 \cdot 0.524}{7.5^2} = 88.889 \cdot 10^{-3} \text{ H;}$$

–inductance of inconsistent position,

$$L_{min} = 107.137 \cdot 10^{-3} - 88.889 \cdot 10^{-3} = 18.249 \cdot 10^{-3} \text{ H.}$$

Next we obtain general formula for inductance levels determination,

$$L = kL_b\gamma, \quad (12)$$

where L_b – basic inductance,

$$L_b = \frac{M_n}{I_n^2};$$

k –factor, what depends on inductance level,

$$k_{max} = \frac{1}{\sqrt{2}},$$

$$k_{min} = \left(\frac{1}{\sqrt{2}} - \frac{2}{m} \right).$$

Now we can design algorithm for inductance levels determination (Fig. 4)

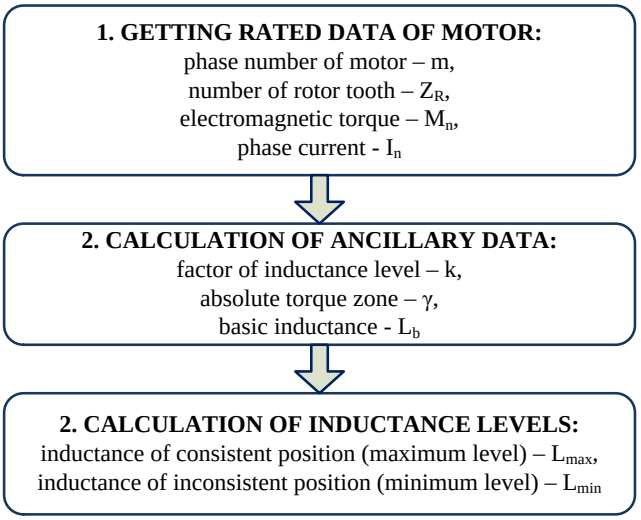


Fig. 4. Short algorithm of declared method.

IV. Conclusion

The general formula for determination of maximum and minimum levels of phase inductance was received. The formula based on rated data of switched reluctance motor. It requires few input data, which are:

–phase number of motor,

–number of rotor tooth,

–electromagnetic torque,

–phase current.

This formula substantially decrease effort of inductance levels determination.

References

[1] Bychkov M. Switched Reluctance Electric Drive: Up-to-date State and Prospect Trends [Online resource] // Journal-guide "Market of Electrical Engineering". – Access mode: http://www.marketelectro.ru/ magazine/readem0207/10 [in Russian]

[2] Bychkova E. Review of Current Market of Frequency Converter for Electric Drive // Industrial Business Yearbook "Zhivaya Elektronika Rossii". – Moscow: OOO "ID Elektronika", 2001. – Vol. 2. – pp. 118–125 [in Russian]

[3] Switched Reluctance Motors [Online textbook] / V. Kuznetsov, V. Kuzmichev. – Access mode: http://elmech.mpei.ac.ru. [in Russian]

[4] Gollandtsev Yu. Switched Reluctance-Jet Motors. – St. Petersburg: GNC RF-CNII "Electropribor", 2003. – 148 p. [in Russian]

[5] Ovchinnikov I. Switched Electric Motors and Drive on their Base (low and average power): Synopsis. – St. Petersburg: KORONA-vek, 2006. – 336 p. [in Russian]

[6] DiRenzo, Michael T. Switched Reluctance Motor Control – Basic Operation and Examples Using the TMS320F240 // Texas Instruments Incorporated. Application Report, SPRA420A. – February, 2000. – 62 p.

[7] Miller, T.J.E. Switched Reluctance Motors and their Control // Magna Physics & Clarendon Press. – Oxford, 1993. – 200 p.

[8] Controlled Alternating Current Drive / B. Firago, L. Pavliachyk. – Minsk: Technoperspectiva, 2006. – 363 p. [in Russian]

[9] Laboratory training session on physics. Part 1. Mechanics / Edited by I. Avenarius and B. Afanasev. – Moscow, State Technical University MADI, 2010. – 107 p. [in Russian]

[10] New in Mechanics / M. Turyishev,V. Shelihov, V. Kuchin et al. – Moscow: OOO"VELMA", 2008. – 17 p. [in Russian]

Dielectric Constant Verification Using COMSOL Multiphysics

[1]*Matúš OCILKA (2nd year),* [2]*Oleksandr PRITCHENKO (4th year)*
Supervisor: [3]*Dobroslav KOVÁČ*

[1,3]Dept. of Theoretical Electrical Engineering and Electrical Measurement, FEI TU of Košice, Slovak Republic
[2]Institute of Electromechanics, Energy Saving and Control Systems, Kremenchuk Mykhailo Ostrohradskyi National University, Ukraine

[1]matus.ocilka@tuke.sk, [2]scenter@kdu.edu.ua, [3]dobroslav.kovac @tuke.sk

***Abstract*—The aim of this paper is experimental measurement of the capacity and estimating relative permittivity of unknown material using laboratory measurement system and comparing results with simulation model created in COMSOL Multiphysics.**

***Keywords*—Capacitance, COMSOL Multiphysics, plate capacitor, relative permittivity,**

I. Introduction

Dielectric constant is a ratio of the electric field strength in vacuum to the electric field strength with medium. The relative permittivity is directly related to the electronic, ionic and orientation of polarization of the material. The first two of these are induced by the applied field, and are caused by displacement of the electrons within the atom, and atoms within the molecule, respectively. The third only exists in polar materials, i.e. those with molecules having a permanent dipole moment. Electronic and ionic polarization are temperature independent, but orientation of polarization, depending on the extent to which the applied field can order the permanent dipoles against the disordering effect of the thermal energy of their environment, varies inversely with absolute temperature. All of these polarization mechanisms can only operate up to a limiting frequency, after which a further frequency increase will result in their disappearance.

The frequency at which these mechanisms drop out is related to the inertia of the moving entities involved. Typically, electronic polarization persists until a frequency of about 10^{16} Hz, ionic polarization until about 10^{13} Hz, while the dispersion for orientation of polarization may lie anywhere within a wide frequency range, say 10^{2}–10^{10} Hz, depending on the material and its temperature. In addition to these polarization mechanisms, the existence of interfacial effects such as macroscopic discontinuities in the material, or blocking at the electrodes, causes the trapping of charge carriers, and such phenomena, as well as the inclusion in the dielectric of impurities giving rise to conducting regions, result in behavior classified under the general heading of Maxwell–Wagner effects. These give rise to an effective polarization and associated loss, the frequency behavior of which is similar to that of orientation of polarization, with a dispersion region which may lie in the region of 1 Hz or lower.

The permittivity of many substances changes not only with frequency and temperature, but also with specimen age and history. Two specimens of nominally the same material may have significantly different permittivity because of different manufacturing processes, different amounts of oxidation, and different inclusions, some of which might have been deliberately introduced, e.g. anti-oxidants. For such reasons, tables of values should be used as an indication of the magnitudes to be expected, and not as a source of precise data which can be repeated by accurate measurements on particular test specimens, except in cases in which the physical and chemical state of both the reference material and the test specimen are very closely specified. The properties of ferroelectric materials depend on so many factors that it is inappropriate to include them in tables of data. Generally, they have permittivity of the order of a thousand, strongly dependent on applied voltage and temperature, and exhibit considerable power loss [2].

In this paper the laboratory measurement system for measuring dielectric constant is introduced. This system is in fact plate capacitor which allows to measure dielectric specimen is in form of thin sheets. Measuring of relative permittivity is made indirectly. First, the capacitance is measured a then using formula (2) the relative permittivity is obtained.

II. Dielectric constant

Dielectric constant also called relative permittivity is defined as a ratio of capacity of capacitor with dielectric to capacity of vacuum capacitor or with capacitor with air gap.

$$\varepsilon_r = \frac{C_x}{C_0} \tag{1}$$

A capacitor filled with a dielectric material has a real capacitance ε_r times greater than would have a capacitor with the same electrodes in vacuum.

A. Dielectric constant of polypropylene

In our case, as a reference specimen thin plastic sheet made of polypropylene (PP) was chosen. PP is the non-polar plastic. In this material there are no polar dipoles present and the application of an electric field does not try to align any dipoles. The electric field does, however, move the electrons slightly in the direction of the electric field to create “electron polarization”, in this case the only movement is that of

electrons and this is effectively instantaneous. These materials tend to have high resistivity and low dielectric constants [3]. Typical values of dielectric constant of PP are shown in Table 1[1].

Values are given for frequencies of 1 kHz, 1 MHz, and 1 GHz; in most cases the dielectric constant at frequencies below 1 kHz does not differ significantly from the value at 1 kHz.

Since the dielectric constant of polymeric material can vary with density, degree of crystallization, and other details (e.g. frequency, temperature, age) of a particular sample as stated before, the values given here should be regarded only as typical or average values [1].

TABLE I
RELATIVE PERMITTIVITY OF POLYPROPYLENE

1kHz	1MHz	1GHz	t[°C]
2.30	2.30	2.30	25.00

B. Calculation of capacitance

Next steps were determining the accuracy and error of measurement system (the reference value was relative permittivity listed in Table 1). Then the measurement of capacity of unknown material was made. Measurement of capacitance is commonly made using (1), but in our case, the measurement system does not allow to set the air gap precisely. The calculation of relative permittivity is made using following formula:

$$C = \varepsilon_r \varepsilon_0 \frac{S}{d} \tag{2}$$

Where C is capacity, ε_r is relative permittivity, ε_0 is permittivity of vacuum, S expresses the area of electrodes and d denotes the distance between the electrodes.

III. MEASUREMENT SYSTEM

A. Geometry of the system

The measurement system is shown in the Fig.1. It consists of two aluminum electrodes. Between the electrodes measured specimen is located. To ensure the distance between the electrodes is the same as the thickness of specimen; electrodes are set between two plates which are tightened with four bolts. Both electrodes should be placed right opposite each other. To ensure the plates and electrodes remains in its place two pins are used. Whole measurement system is described by following parameters. a_w is width of plate[mm], a_h is height of plate [mm], b_w is width of electrode [mm], b_h expresses the height of electrode [mm], b_t is thickness of electrode [mm], and t is thickness of specimen[mm].

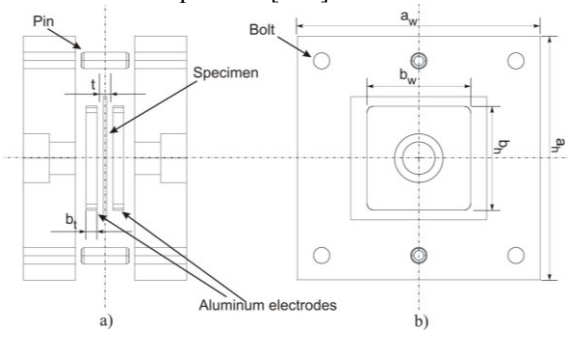


Fig. 1. Measurement system

B. Model of system - plate capacitor

Model of measurement system (plate capacitor) is created in COMSOL Multiphysics software in AC/DC module (Fig.2). COMSOL Multiphysics is software package that solves systems of coupled three-dimensional partial differential equations. This software can be used to model different physical phenomena including electromagnetics. It uses the finite element method of solution and can model the complex problems, geometries or material properties in 2D or 3D representation for better understanding some phenomena. This model is simplified and consists only of electrodes and specimen that are placed in air domain.

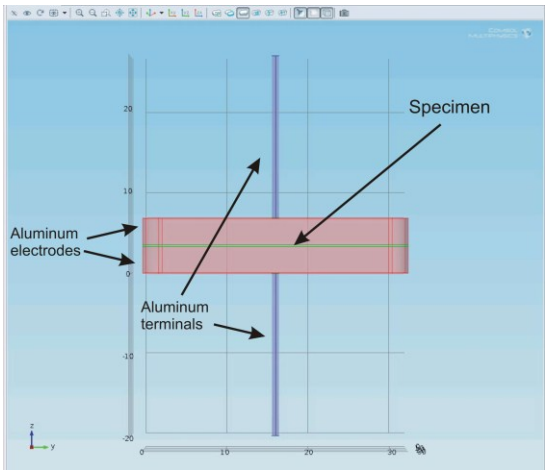


Fig. 2. Model of measurement system in COMSOL Multiphysics

IV. RESULTS

A. Simulation results

For simulation the model shown in Fig.2 was used. As the material of electrodes aluminum and as the dielectric material PP with relative permittivity $\varepsilon_r = 2.3$ was selected as stated in Table 1. Problem was solved as electrostatic problem.

Very important parameter of simulation is a mesh. Proper size and shape of mesh elements influences the accuracy of solution; the finer the mesh elements, the better the solution accuracy. Of course, very fine mesh can be a problem and computer cannot solve such model. It is very important to find optimal number of elements. In our case the tetrahedral mesh of 691 737 elements was created. Parameters used in simulation:

Height of electrode/specimen:	b_h = 32mm
Width of electrode/specimen:	b_w = 32mm
Thickness of electrodes:	b_t = 3.3mm
Thickness of specimen:	t =0.1mm
Radius of round corners:	r = 2mm
Area of electrodes/specimen:	S=1020.57mm^2
Physics	electrostatics
Relative permittivity (PP):	ε_r = 2.3
Number of mesh elements:	691 737
Applied voltage:	V_0=1V

The simulation results are shown in Fig.3. The simulation took 48s and capacitance calculated by software is $C_{simPP} = 210.6$ pF. The capacitance using (2) is $C_{calcPP} = 207.8$ pF.

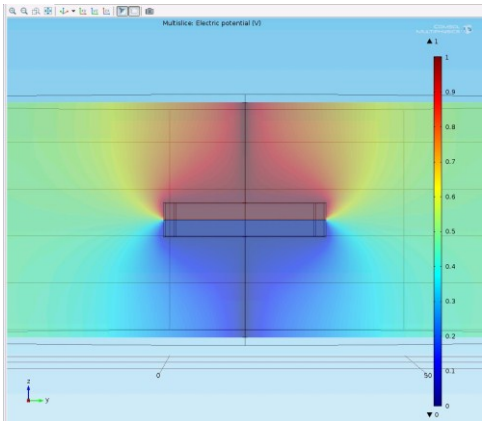


Fig. 3. Electric potential [V]

B. Measurement results with PP specimen

The measurement system is depicted in Fig.4 and measurements were made using LCR bringe Hameg HM8118 with basic accuracy 0.05% (Fig.5). The thickness of used PP sheet was 0.05mm and in measurement two pieces of specimen were used.

TABLE II
CAPACITOR WITH PP DIELECTRIC

	20Hz	60Hz	t[°C]
1.	1.94E-10	1.93E-10	20
2.	1.90E-10	1.89E-10	20
3.	1.94E-10	1.92E-10	20
4.	1.90E-10	1.89E-10	20
5.	1.91E-10	1.90E-10	20
6.	1.90E-10	1.88E-10	20
7.	1.91E-10	1.90E-10	20
8.	1.90E-10	1.89E-10	20
9.	1.90E-10	1.90E-10	20
10.	1.90E-10	1.90E-10	20
Average:	**1.91E-10**	**1.9E-10**	

For better accuracy 10 measurements were taken, always with new specimen. The measurement was made at 20Hz and 60Hz. The values of obtained capacitance [F] are listed in Table 2.

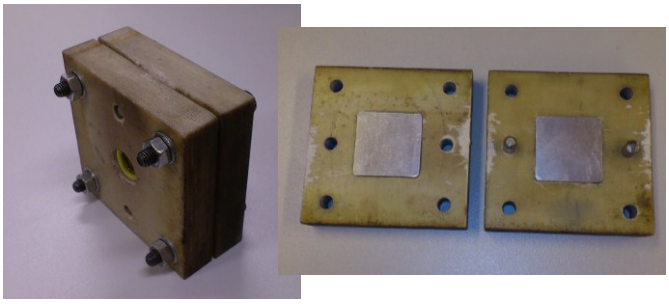


Fig. 4. Measurement system

Fig. 5. LCR measurement bridge HM8118

Relative permittivity of PP for every measurement is then calculated using (2) and shown in Table 3. Average value of permittivity of measured PP specimen at 20 Hz is ε_r=2.114.

TABLE III
RELATIVE PERMITTIVITY OF PP

	20[Hz]	60[Hz]
1.	2.141	2.131
2.	2.106	2.095
3.	2.144	2.128
4.	2.105	2.093
5.	2.116	2.103
6.	2.097	2.075
7.	2.114	2.100
8.	2.105	2.090
9.	2.106	2.103
10.	2.105	2.103
Average:	**2.114**	**2.102**

C. Measurement results with unknown specimen

The thickness of used specimen was 0,155 mm. Again, 10 measurements were taken, always with new specimen at frequencies 20Hz and 60Hz. The values of obtained capacitance [F] are listed in Table 4.

TABLE IV
CAPACITOR WITH UNKNOWN DIELECTRIC

	20	60	t[°C]
1.	1.75E-10	1.73E-10	20
2.	1.74E-10	1.73E-10	20
3.	1.73E-10	1.72E-10	20
4.	1.84E-10	1.81E-10	20
5.	1.84E-10	1.82E-10	20
6.	1.73E-10	1.71E-10	20
7.	1.74E-10	1.72E-10	20
8.	1.78E-10	1.76E-10	20
9.	1.76E-10	1.74E-10	20
10.	1.72E-10	1.71E-10	20
Average:	**1.76E-10**	**1.74E-10**	

Relative permittivity of unknown material for measured capacitance is shown in Table 5.

TABLE V
CAPACITOR WITH UNKNOWN DIELECTRIC

	20[Hz]	60[Hz]
1.	2.995	2.971
2.	2.987	2.961
3.	2.974	2.943
4.	3.153	3.110
5.	3.151	3.120
6.	2.967	2.940
7.	2.981	2.954
8.	3.057	3.021
9.	3.015	2.985
10.	2.954	2.925
Average:	**3.023**	**2.993**

For simulation model of capacitor filled with unknown dielectric the model depicted in Fig.2 was used. The parameters of the model remain the same, except for the parameters listed below:

Thickness of specimen:	t =0.155mm
Relative permittivity:	ε_r = 3.023
Number of mesh elements:	756 399

For the value of relative permittivity of dielectric material used in simulation the average value at 20Hz was set. The capacitance calculated by software is C_{simU}=178.6pF.

D.Results comparison

In the Table 6 the results of analytic, simulation and experimental method are summarized. Measured permittivity is used in analytic and simulation approach and compared with experimental measurement.

TABLE VI
COMPARISON BETWEEN ANALYTIC METHOD, SIMULATION AND MEASUREMENT

	Analytic	Simulation (Electrostatics)	Measurement Average (20Hz)
PP	1.91E-10	1,94E-10	1.91E-10
ε_r = 2.114			
Unknown	1.76E-10	1.79E-10	1.76E-10
ε_r = 3.023			

The experimental results show that relative permittivity of PP differs (8.1%) from permittivity stated in literature. This can be caused by various factors including temperature, frequency, density or age of particular specimen. There are other factors which can also influence the accuracy of measurement, e.g. accuracy of LCR Bridge or accuracy during fabrication of measurement system (poor contact of electrodes or electrodes can be slightly shifted from each other causing lesser contact area). We can assume that the average value of relative permittivity (in this case) and the capacity of capacitor filled with this unknown material may differ from real value with relative deviation of 8,1%.

V.CONCLUSION

In this paper the measurement and calculation of relative permittivity was presented. For our purpose the measurement system that consists of two aluminum electrodes was built. Between electrodes the dielectric specimen was placed. This system represents the plate capacitor which capacity can be easily calculated whether analytic or simulation method. The simulation of plate capacitor was made using COMSOL Multiphysics software which solved the electric fields and calculated the capacitance of such capacitor. First, the measurement of relative permittivity for known material – polypropylene was made. Then the measurement of capacity of unknown material and calculation of relative permittivity was made. In simulation model the average values from both measurements were used for comparison results with experimental approach. This system allows measuring of capacitance of plate capacitor filled with thin dielectric sheet with good accuracy. Together with finite elements method of solution this system can be used in educational process for students to better understand the phenomenon of electromagnetism and for practical experiments.

ACKNOWLEDGMENT

The paper has been prepared under support of Slovak grant projects KEGA No. 005TUKE-4/2012.

REFERENCES

[1] David R. Lide, “CRC Handbook of Chemistry and Physics,”88th edition, 2007-2008

[2] National Physical Laboratory Kaye and Laby, “Tables of Physical and Chemical Constants,” 2013

[3] Zeus Technical Whitepaper, ”Dielectric properties of polymers,” Zeus Industrial Products, Inc.,2005

[4] D. Kováč, M. Ocilka, M. Vansáč, “Calculating of inductance of spiral coil using comsol multiphysics,” In: Electromechanical and energy systems, modeling and optimization methods: conference proceedings: the 10th international conference of students and young researchers: March 28-29, 2012, Kremenchuk. - Kremenchuk : Mykhailo Ostrohradskyi National University, 2012 P. 89-90. - ISSN 2079-5106

[5] M. Ocilka, O. Kravetz, O. Shutka, “Calculation of inductance of double-layer planar spiral coil using COMSOL Multiphysics,” 1 elektronický optický disk (CD-ROM), In: SCYR 2012 : proceedings from conference : 12th Scientific Conference of Young Researchers : May 15th, 2012, Herľany, Slovakia. - Košice: TU , 2012 S. 247-249. - ISBN 978-80-553-0943-9

Dielectric Relaxation Spectroscopy of XLPE Cables in Frequency Domain

[1]*Martin GERMAN-SOBEK (2st year),* [2]*Jozef KIRÁLY (2st year)*
Supervisor: [3]*Roman CIMBALA*

[1]Dept. of Electric Power Engineering, FEI TU of Košice, Slovak Republic

[1]martin.german-sobek@tuke.sk, [2]jozef.király@tuke.sk, [3]roman.cimbala@tuke.sk

***Abstract*— Cross-linked polyethylene (XLPE) is widely used as modern insulation for high voltage cables. The thermal aging during operation changes its quality and electrical parameters. The article deals with the measuring of dielectric parameters of XLPE cable sample. The measurement was carried out by the method of dielectric relaxation spectroscopy (DRS) in frequency domain. It was observed a change of parameters in consequence of additional aging.**

***Keywords*— XLPE cable, capacitance, dissipation factor, thermal aging, dielectric relaxation spectroscopy**

I. Introduction

In the present, for high-voltage electrical equipment is required greater security and reliability. These requirements relate mainly to the insulation system, which is an important and sensitive part of any electrical equipment. Power cables are very important and sensitive devices in the power system, and they play an important role in the safety of the power load and reliable transmission of electricity.

Nowadays, the majority of power cables are insulated with polymeric materials. Cross-linked polyethylene (XLPE), as the main polymeric insulation, is widely used as electrical insulation material for high-voltage distribution power cables. This insulation material provides excellent physical, chemical and electrical properties. Given that these cables may be exposed to several kA current and hundreds of kV voltage, they are critical parts of the transmission infrastructure and is expected their high resistance against possible failures during their lifetime [1]. Damage of insulation can lead to equipment failure and other disorders.

The cables are permanently exposed to thermal aging during operation. It may cause change in dielectric parameters of cables and also irreversible damage of cable insulation. Process of aging of insulation is the most acting on the quality of insulation and it is a phenomenon that is essentially cannot be affected. The insulation degradation is inevitable during the operation and the failure rate of XLPE cable increases with the service time [2]. It is necessary to assess degradation and insulating state of cables.

There are several methods for detection of insulating state of XLPE insulation. One of the methods is measurements of dielectric parameters of insulation using the dielectric relaxation spectroscopy (DRS).

II. About Dielectric Relaxation Spectroscopy

This method is one of the non-destructive measurement methods. DRS uses polarization as the response of the sample on a time-dependent electric field. For electrolyte solutions, polarization essentially originates from the oriented fluctuations of permanent dipoles (solvent molecules, ion pairs), from intramolecular polarizability and from ion motion. It can be investigated either in the time domain or as a function of the frequency of a harmonic field. Thus, principle of this method is based on examination of molecular dynamics of polarized and polar materials. DRS is widely applied in the characterization of ion-conducting solids and polymers. [3], [4]

Generally speaking, the behaviour of the dielectric material is characterized DC conductivity, dielectric response function and high-frequency component of the relative permittivity. [4]

In dielectric spectroscopy in the frequency domain, the polarization phenomena follow the alternating electric field. DRS is thus focused on the measurement of the frequency dependence of real and imaginary components of the impedance of the investigated samples. DRS evaluates the dielectric response function in the frequency domain using the dielectric dissipation loss factor $tan\delta$ and complex capacitance $C(\omega)$. For an accurate evaluation of the properties of the investigated dielectrics is a sometimes necessary using appropriate transformation method to convert measured values from time to frequency domain, e.g. using the Laplace, or Fourier transform, or Hamon approximation. This article deals only with the measurement of frequency response. [4]

The total current flowing across the material when exposed to voltage $U(\omega)$ can be expressed as:

$$I(\omega) = i\omega C_0 \left[\varepsilon_\infty + \chi'(\omega) - i\left(\frac{\sigma_0}{\varepsilon_0 \omega} + \chi''(\omega)\right)\right] U(\omega) \qquad (1)$$

Complex electrical induction $\underline{D}(\omega)$ is proportional to the complex dielectric permittivity $\underline{\varepsilon}(\omega)$ according to the relation [4]:

$$\underline{D}(\omega) = \varepsilon_0 \underline{\varepsilon}(\omega)\, \underline{E}(\omega) = \varepsilon_0 \left[1 + \chi'(\omega) - \underline{i}\, \chi''(\omega)\right] \underline{E}(\omega) \quad (2)$$

where

$$\underline{\varepsilon}(\omega) = \varepsilon'(\omega) - \underline{i}\,\varepsilon''(\omega) = (1 + \chi'(\omega)) - \underline{i}\,\chi''(\omega) \qquad (3)$$

Actual measurements of this dielectric response in the frequency domain are difficult to perform, if the frequency range becomes very large. The measured relative dielectric permittivity $\underline{\tilde{\varepsilon}}_r(\omega)$ is defined from the following relation [4]:

$$\underline{j}(\omega) = i\omega\varepsilon_0\underline{\tilde{\varepsilon}}_r(\omega)\,\underline{E}(\omega) \qquad (4)$$

Therefore

$$\underline{\tilde{\varepsilon}}_r(\omega) = \varepsilon'_r(\omega) - i\left[\varepsilon''_r(\omega) + \frac{\sigma_0}{\varepsilon_0\omega}\right] = 1 + \chi'(\omega) - i[\chi''(\omega) + \sigma_0/\varepsilon_0\omega] \qquad (5)$$

The dielectric dissipation factor:

$$tan\delta(\omega) = \frac{\varepsilon''_r(\omega)+\sigma_0/\varepsilon_0\omega}{\varepsilon'_r(\omega)} \qquad (6)$$

The real part of (5) represents the capacitance of a test object, whereas the imaginary part represents the losses. Both quantities depend on frequency. It should be noted that all dielectric quantities are more or less dependent on temperature. Any comparison or measurement of these quantities must take this into account. Increased interfacial polarization also produces the increase in dissipation factor, mainly in the low and very low frequency range. [4]

Assuming that the real part of the permittivity (ε') of XLPE to remain reasonably constant over the frequency range studied and not to be particularly influenced by electric field strength, the dielectric loss ($tan\delta$) would therefore tend to have the same trends with frequency as the imaginary component of the permittivity (ε'') since $tan\delta = \varepsilon''/\varepsilon'$. [1]

Measurements in the frequency domain need voltage sources of variable frequencies and, for applications related to HV power equipment, output voltages up to at least some hundreds of volts. Such measurements become quite lengthy if very low frequencies are considered.

III. XLPE UNDER OPERATION CONDITION

During the operation, an insulation system is subjected to one or more stress that causes irreversible changes of insulating material properties with time. This process is called aging and ends when the insulation is no more able to withstand the applied stress. The stresses most commonly applied in operation are electric field (due to voltage) and temperature (due to loss), but also other stresses, such as mechanical stresses (vibration) and environmental stresses (pollution, humidity) can be present. These stresses are the primary initiators for the degradation of insulation. Interaction various factors together may significantly speed up the degradation processes. Process of aging of insulation is the most acting on the parameters and quality of insulation. [5]

Aging of XLPE cables during operation is related to the temperature of the insulation. All XLPE cables contain antioxidants which protect the XLPE from oxidation during the service life of the cable. The rate at which the antioxidant is used up is dependent on temperature. The normal maximum operating temperature of XLPE cables is 90°C. Increasing the operating temperature of the cables will increase the rate at which the anti-oxidant is used up and hence reduce the service life. Small increase in temperature has a significant impact on the aging of the XLPE. [6]

Tests have shown that XLPE cables can operate at a temperature of 105°C for a limited time without significantly reducing the service life of cables. At temperatures in excess of 105°C deformation of XLPE readily occurs, particularly at positions where the insulation is under mechanical stress. The maximum overload temperature of XLPE is limited to 105°C. [6], [7]

Stability of microstructure and composition of the insulating material is changing due to degradation processes. These changes result to changing behaviour of insulation material from view of polarization processes. Aging in a polymer changes the electrical, physical, mechanical and morphological properties of the insulation. All these properties are influencing the dielectric parameters and characteristics of insulation. Many methods have been proposed to evaluate the properties of XLPE. [8]

IV. EXPERIMENT

Diagnostic measurement was performed by the method of dielectric relaxation spectroscopy in frequency domain. Using the precision LCR meter Agilent E4980A were recorded changes of capacitance and dielectric dissipation loss factor depending on frequency for one sample of HV cable.

The sample was aged and degraded cable of operationally unknown technical condition. It was a power cable with an aluminium core and XLPE insulation. Sample of cable was approximately 25 cm long and protective cable jacket and semiconducting layer have been removed (2 cm at both ends) and shielding taken out.

The frequency range of measurement was 20 Hz to 2 MHz. The Increase of frequency was decimal. The measurement was repeated after an accelerated thermal aging of sample at 90 °C for 72 hours in an air oven.

Aim measuring has been comparison of measured frequency dependence of capacitance and dissipation factor of cable sample before and after aging process.

Circuit configuration for measurement is shown in Figure 1, where C_X is measured sample of cable.

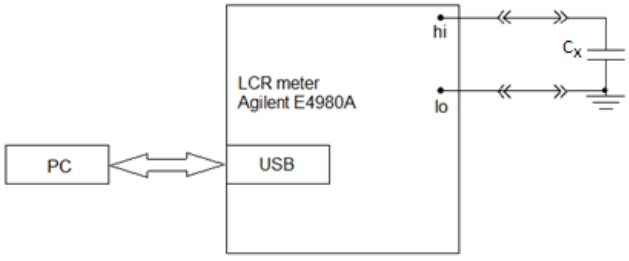


Fig. 1. Schematic diagram of measuring circuit

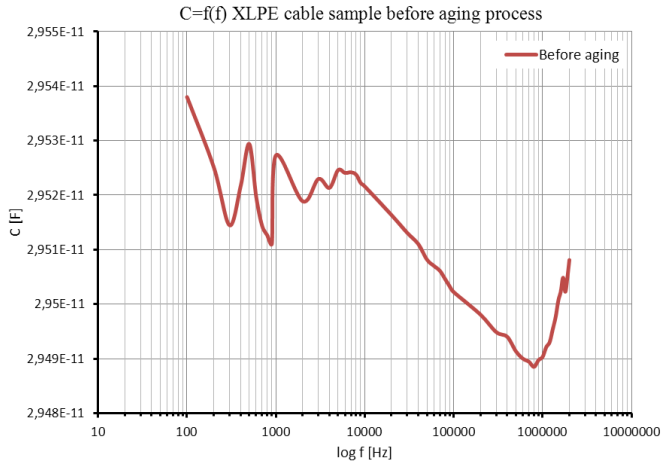


Fig. 2. Frequency dependence of capacitance of sample before aging process

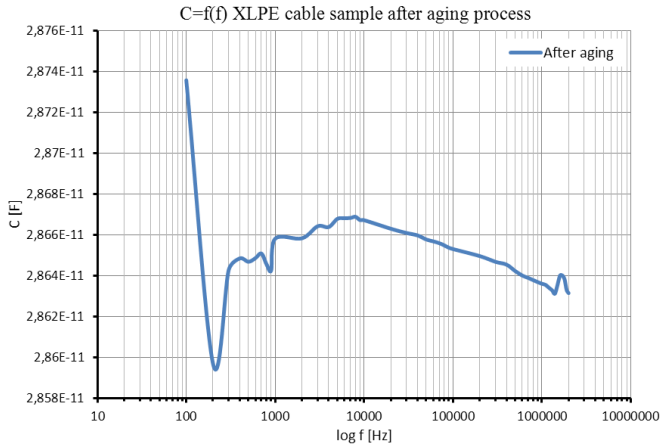


Fig. 3. Frequency dependence of capacitance of sample after aging process

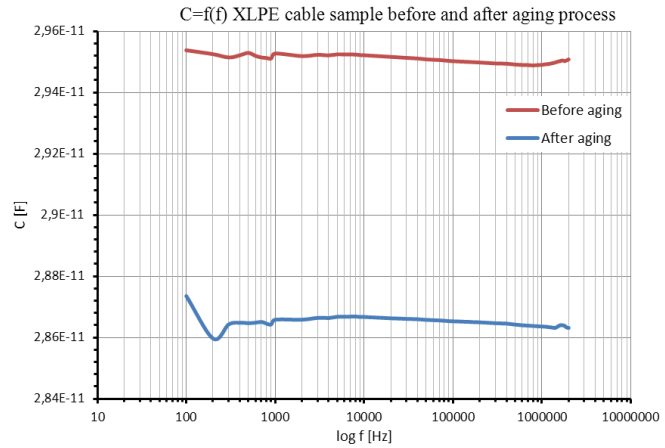


Fig. 4. Comparison of frequency dependence of capacitance before and after aging process

Figure 2 and Figure 3 shows the frequency dependence of capacitance before and after aging, respectively. In Figure 4 is shown comparison of frequency dependence of capacitance before and after aging process. Figure 5 and Figure 6 shows the frequency dependence of dissipation factor before and after aging, respectively. In Figure 7 is shown comparison of frequency dependence of dissipation factor before and after aging process. Due to the large fluctuations in the measured values at low frequencies from 20 Hz to 100 Hz, the measured values are not within the zone of termination capable and therefore are not shown.

From dependencies of capacity can be seen decrease of the values of capacitance after aging process. The change of capacitance is very small, order of 10^{-11} farad. The resulting dependence of capacitance could be affected by disturbances and surroundings, since the sample size was small with a very small capacitance. Such measurement is sensitive to disturbances and parasitic capacitance.

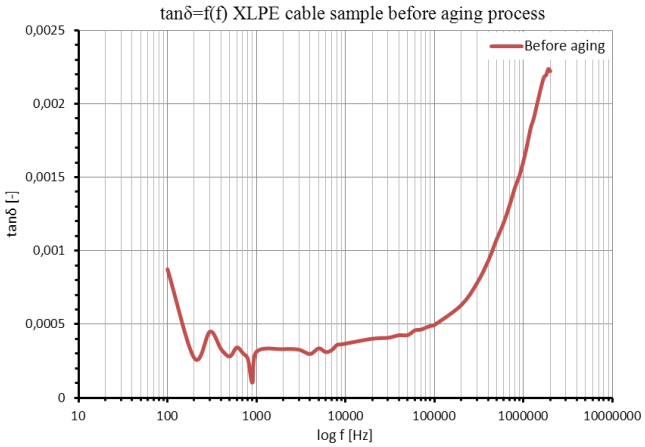


Fig. 5. Frequency dependence of dissipation factor of sample before aging process

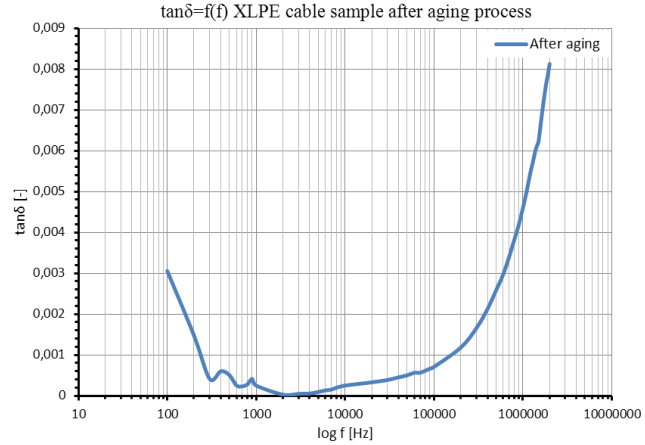


Fig. 6. Frequency dependence of dissipation factor of sample after aging process

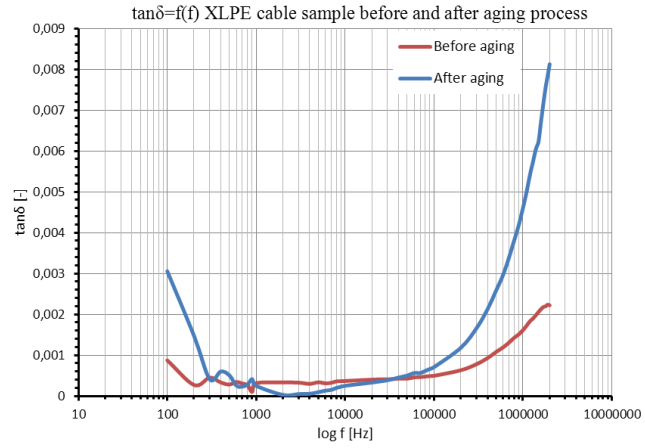


Fig. 7. Comparison of frequency dependence of dissipation factor before and after aging process

From dependencies of dissipation factor can be seen a change of the values of dissipation factor after aging process. The change of dissipation factor is significant at low

frequency up to 200 Hz and at high frequency from 100 kHz where there was an increase of values. In frequency range between 1 kHz and 10 kHz was decrease of values. According to [1] above 100 Hz it is dominated by the series resistance of the semiconductors and behaviour of dissipation factor is dominated by the series resistance of the semiconducting layers.

V. Conclusion

The dielectric properties and electrical parameters of XLPE insulation are dependent on a number of factors that affect to the insulation during operation. It is known that thermal aging of insulation influences on the parameters and quality of insulation. Using DRS in frequency domain can be evaluated change of properties of investigated insulation material after aging and degradation processes during operation.

Aim of this paper was to comparison and evaluation of measured frequency dependencies of capacitance and dissipation factor of operationally aged sample of XLPE cable before and after the additional degradation. For measurement was chosen method of DRS. The showed dependencies of capacitance and dissipation factor show that aging process change dielectric properties of sample of XLPE cable. Such changes can be related to structural changes in the polyethylene morphology. The measurement confirmed the suitability of DRS for investigate cable insulation system. It is not possible exclude the surrounding of a disturbance during the measurement, which could affect the measured results. It is important to continue to investigate the process of aging and their impact on the insulation system with use modern and suitable diagnostic methods.

Acknowledgment

We support research activities in Slovakia / Project is cofinanced from EU funds. This paper was developed within the Project "Centrum excelentnosti integrovaného výskumu a využitia progresívnych materiálov a technológií v oblasti automobilovej elektroniky", ITMS 26220120055.

This work was supported by scientific Grant Agency of the ministry of Education of the Slovak Republic project VEGA No. 1/0487/12.

We support research activities in Slovakia / Project is cofinanced from EU funds.

References

[1] Fothergrill, J. C., et al.: "The Measurement of Very Low Conductivity and Dielectric Loss in XLPE Cables," *IEE Transaction on Dielectrics Electrical Insulation*, Vol. 15, No. 5, 2011, pp. 1544-1553.

[2] Shuvalov, M. et al.: Analysis of Water Trees in Power Cable Polymeric Insulation. In: Journal Applied Polymer Science, Vol.88, 2003, pp. 1543-1549.

[3] Buchner R.: "Dielectric spectroscopy of solutions". In: Novel Approaches to the Structure and Dynamics of Liquids: Experiments, Theories and Simulations, 2004, pp. 265–288.

[4] Zaengl, W. S.: "Dielectric Spectroscopy in Time Domain for HPV Power Equipment, Part I: Theoretical Consideration," *IEE Transaction on Dielectrics Electrical Insulation*, Vol. 19, No. 5, 2003, pp. 5-19.

[5] Rawangpai, A., et al.: Artificial Accelerated Ageing Test of 22 kV XLPE Cable for Distribution System Applications in Thailand. In: World Academy of Science, Engineering and Technology 65 2010, pp. 220-225.

[6] Ageing of XLPE Compounds, Brochure, General Cable Australia Pty Ltd.

[7] Metwally, I. A.: The Evolution of Medium Voltage Power Cables. In: Potentials, IEEE , Vol.31, No.3, 2012, pp. 20-25.

[8] Hoff, G., Kranz, H. G.: "Interpretation of Dielectric Response Measurement Data from Service Aged XLPE-Cables," *IEEE 7th International Conference on Solid Dielectrics*, Eindhoven, June 25-29, 2001, pp. 381-384.

[9] Wald, D., Hampton, N.: How much does studying Polyethylene tell us about XLPE? In: Electrical Insulation (ISEI), Conference Record of the 2012 IEEE International Symposium, 2012 pp.250-254.

[10] Fei Liu et al.: Insulation ageing diagnosis of XLPE power cables under service conditions. In: Condition Monitoring and Diagnosis (CMD), 2012 International Conference, 2012, pp.647-650.

[11] Muhr, M., Schwarz, R., Jaufer, S.: "Electrical Measurements as Diagnostic Tool for HV-Insulations," Graz, 2005, pp. 1-11.

Experimental investigations of working modes of power thyristor keys in semiconductor converters

[1]*Viacheslav MELNYKOV,* [2]*Ján PERDUĽAK*
Supervisor: [3]*Andrii KALINOV*

[1,3] Institute of Electromechanics, Energy Saving and Control Systems Kremenchuk Mykhailo Ostrohradskyi National University, Ukraine
[2]Dept. of Theoretical Electrical Engineering and Electrical Measurement, FEI TU of Košice, Slovak Republic

[1]scenter@kdu.edu.ua, [2]jan.perdulak@tuke.sk

***Abstract* – In this paper experimental investigation of IGBT usage as a switching element of semiconductor converter is presented. Investigations concerning influence of additional circuits of transistor working point changing under switching on energy losses were made. An evaluation of overvoltage guard circuits effectiveness and their influence on IGBT switching was done.**

***Keywords* – Semiconductor key, dynamic losses, switching speed, guard circuits.**

I. INTRODUCTION

A progress of modern semiconductor technology is characterized by usage of power keys with high switching speed, such as MIS transistors and IGBT. In such power keys a transition from open to closed position and vice versa occurs in a short time interval which, as a rule, is not comparable to steady state of the key. Transistor instantaneous currents and voltages determine working point position on a space of current-voltage curves [1]. Consecutive time transition from one working point to another is estimated by working point trajectory. This trajectory depends on electric circuit and control signals parameters.

Using IGBT as transistor switching element it should be ensured that its working point always is inside the secure working area (SWA) without crossing border of this area, because even short-time crossing leads to transistor going out to breakdown area [2]. In order to protect IGBT from breakdown, usually its moving trajectory should be formed. In this case additional guard circuits should be formed.

Usage of additional circuits for forming the transistor working point in switching modes leads to decreasing losses and avoid parasitic current and voltage jumps [1].

II. AN EVALUATION OF ENERGY LOSSES IN IGBT SWITCHING

Nowadays power drivers are widely used for IGBT control [3]. Drivers' connection to transistor input circuits boils down to solving of constructive questions of mounting parasitic inductances minimization and also to usage of additional schematic solutions for switching speed control of power key and excluding the influence of feedback between transistor input and output circuits.

Transistor switching speed control could be done using series resistor between driver input node and key output circuit.

For keys with high switching speed a usage of limiting resistors under high load current amplitudes is obligatory for safety work [1, 4]. Although, switching speed limitation leads to increasing of dynamic power losses.

In general case work of semiconductor keys specified by losses, which could be the following types: transient losses in turning-on and turning-off modes, in activated or deactivated modes, in input control circuit [1]. Switching processes influence significantly on energy losses and stable work of key elements. In transients transistor appears in high voltage state with significant direct current which leads to significant energy losses. If switching time of key element is not limited, a part of switching losses in general energy losses could become significant.

In order to analyze energy losses of power transistor, a direct current impulse thru the key is used (fig. 1). According fig. 1, current loads are characterized by the following time parameters: $t_P = t_3 - t_2$ – direct current time; $t_R = t_2 - t_1$ – direct current rise time; $t_F = t_4 - t_3$ – direct current fall time.

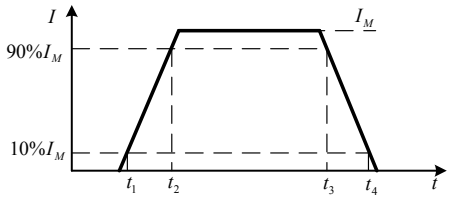


Fig. 1. Shape of a direct current impulse thru transistor.

Losses energy for turning-on (E_{ON}) and turning-off (E_{OFF}) modes could be calculated according to the following expressios:

$$e_{ON} = \int_{t_1}^{t_2} i(t)u(t)dt\,, \quad e_{OFF} = \int_{t_3}^{t_4} i(t)u(t)dt\,. \qquad (1)$$

Presented expressions are typical for energy losses determination for one switch of transistor key. For calculation of average energy losses in key for one hour the next expression could be used:

$$E_{ON} = e_{ON} \cdot f \cdot 3600, \quad E_{OFF} = e_{OFF} \cdot f \cdot 3600, \qquad (2)$$

where f – transistor key switching frequency.

A usage of limiting resistors for connection of driver output circuit to IGBT input circuit is a simple and effective solution. However, in converters with electrical motors the power key switching-on time should be increased in order to limit the current jump in starting mode and mode of decreasing the rotational speed. In order to limit the voltage jumps in switching-off modes, the transistor closing speed should be limited. In this key both unidirectional additional circuits (fig. 2) or drivers with split control channels of switching-on and switching-off [4] could be used.

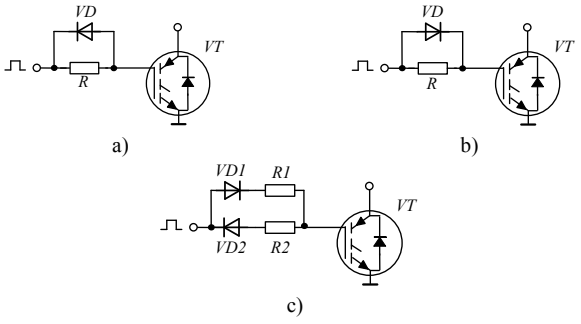


Fig. 2. A transistor switching speed limitation: for turning-on mode (a), for turning-off mode (b), for split of turning-on and turning-off modes (c).

III. Experimental researches on the influence of additional circuits

Experimental researches were done for investigations of the influence of the additional circuits of key switching trajectory formation for schemes with different parameters and for transistor key work with active and active-inductive load. In the fig. 3 current, voltage and losses power signal are presented for transistor key turning-on and turning-off modes (connection circuit is presented in fig. 2, c) for usage resistors of 24 kOhm (fig. 3, a) and 400 kOhm (fig. 3, b). In fig. 4 the dependences of energy dynamic losses for IGBT work on active (fig. 4, a) and active-inductive load (fig. 4, b) for different types of additional circuits.

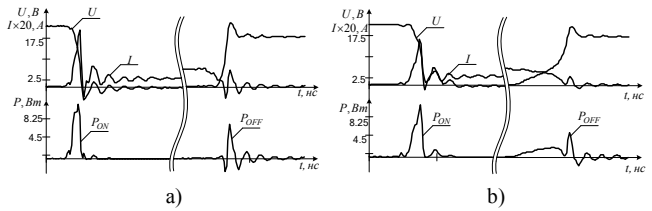


Fig. 3. Current, voltage and losses power signals for transistor key turning-on and turning-off modes with resistances of 24 kOhm (a) and 400 kOhm (b).

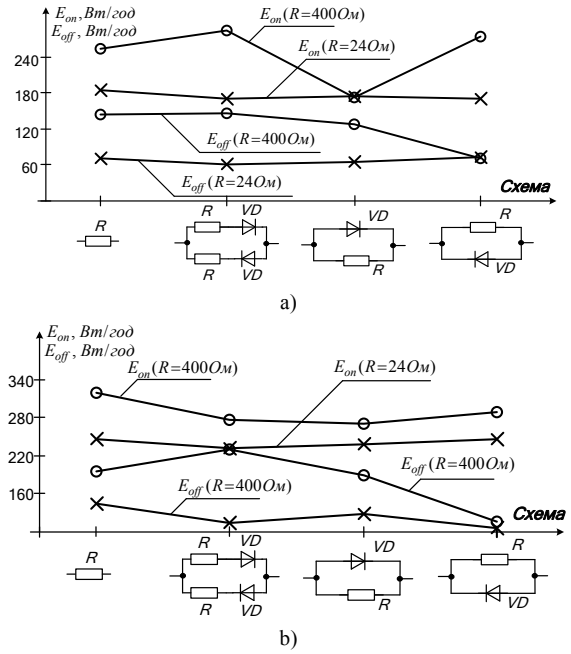


Fig. 4. The influence of the additional circuits of IGBT working point formation on dynamic energy losses for working modes with active (a) and active-inductive (b) load

IV. IGBT safety circuits

Power semiconductor keys inclined to breakdown possibility, and their working characteristics determined by limit exploitation parameters. All parameters of thyristor limit working modes specified by evolution of the one of the breakdown types: electrical – as a result of high voltage, or thermal – as a result of current overheating.

In order to decrease possibility of appearance one of the key breakdown modes, the requirement for semiconductor energy converter designing is accurate construction, compact positioning of power elements and minimization of electrical connections between them [4]. These conditions could be described by the following: if scheme of impulse converter (fig. 5) contains input capacitor C_{in} connected to another part of scheme with long wires, this connection could be presented by parasitic inductance. While current i_{VD} flows, the energy will be accumulating in capacitor. When transistor VT is shutting, collector current stop flowing, but energy which was accumulated in parasitic inductance will try to support this current, that's why voltage jump between collector and emitter appears and it makes some "addition" to supply voltage. A current start flowing thru diode VD, transistor emitter is connected to common scheme wire and "addition" sums to input voltage. The higher speed of collector current drop is, the biggest "addition" could be given and more dangerous key transistor working mode is.

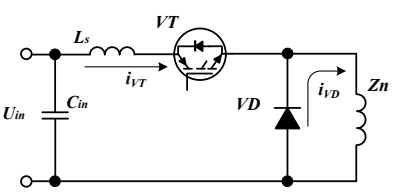


Fig. 5. Impulse converter with impulse inductance.

Thus, key input voltage could be easily taking into account but forecasting of overvoltage influence is still difficult task. Thus, in order to increase converter reliability safety elements should be included into the circuit. Nowadays except for classical overvoltage guard methods, such as *RC* and *RCD* circuits, the varistors and suppressors also became popular [1]. Usage of additional safety elements allows exclude parasitic voltage and current jumps, but they significantly influence on dynamic energy losses.

V. Experimental researches on IGBT overvoltage safety schemes

Experimental investigations of IGBT overvoltage safety schemes usage were done for impulse converter work with active and active-inductive load. As safety circuits schemes were chosen *RC* and *RCD* circuits, varistors and suppressors. Evaluation of their effectiveness were made basing on the following criteria: relative overvoltage in IGBT turning-on and turning-off mode (Δu_{on}, Δu_{off}), turning-on and turning-off time (t_{on}, t_{off}), energy dynamic losses for converter one hour work (E_{on}, E_{off}).

In connecting wires of presented impulse energy converter (fig. 5) it could be several reasons of parasitic inductances appearance. Accounting this, experimental investigations were done for the following working modes: parasitic inductance is absent (mode #1), parasitic inductance is placed before transistor collector (mode #2), parasitic inductance is placed after transistor emitter and before backward diode (mode #3) and parasitic inductances are placed before collector and after emitter of transistor (mode #4). Overvoltages for IGBT turning-on and turning-off modes for mentioned working modes without usage of safety circuits are presented in fig. 6.

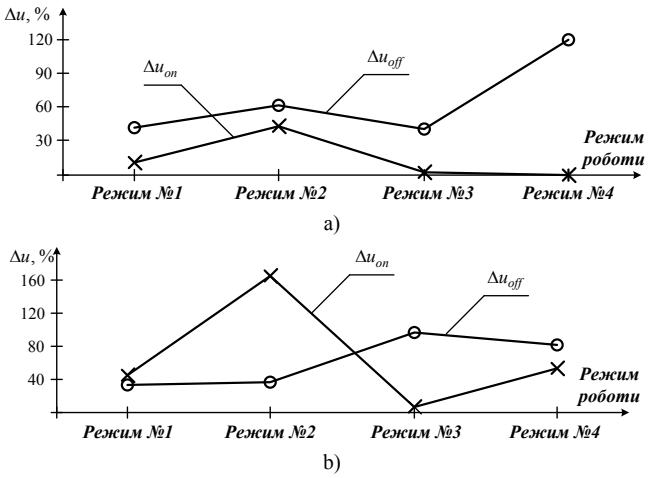


Fig. 6. Overvoltages in IGBT switching modes for work with active (a) and active-inductive (b) load.

In fig. 7–9 the effectiveness of usage the different types of the IGBT safety circuits for energy converter work with active load is presented. It should be mentioned, that overvoltages in IGBT turning-on modes do not depend on safety circuit scheme and could be determined just parasitic inductance position. For all presented characteristics the following keys are used: mode # 1 – ✕✕✕, mode # 2 – ▭▭▭, mode # 3 – ◇◇◇, pmode # 4 – ⊖⊖⊖.

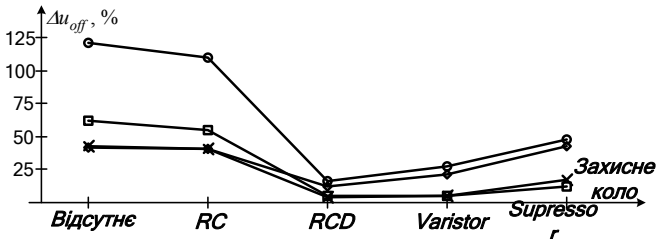


Fig. 7. Overvoltage dissipation for IGBT turning-off mode.

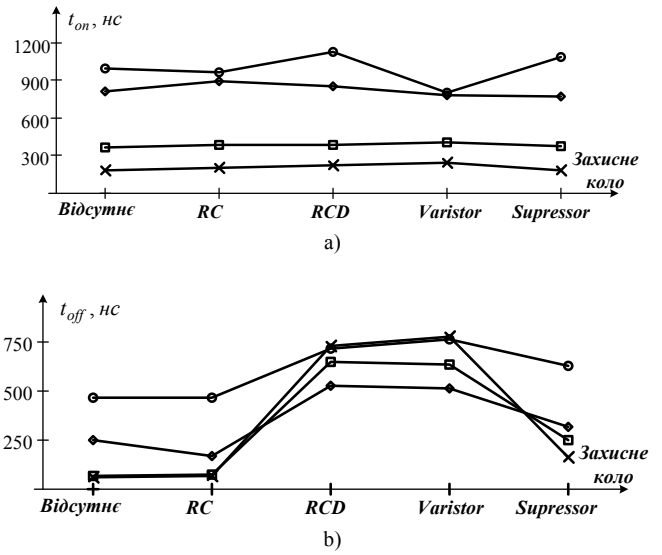


Fig. 8. IGBT turning-on (a) and turning-off (b) time.

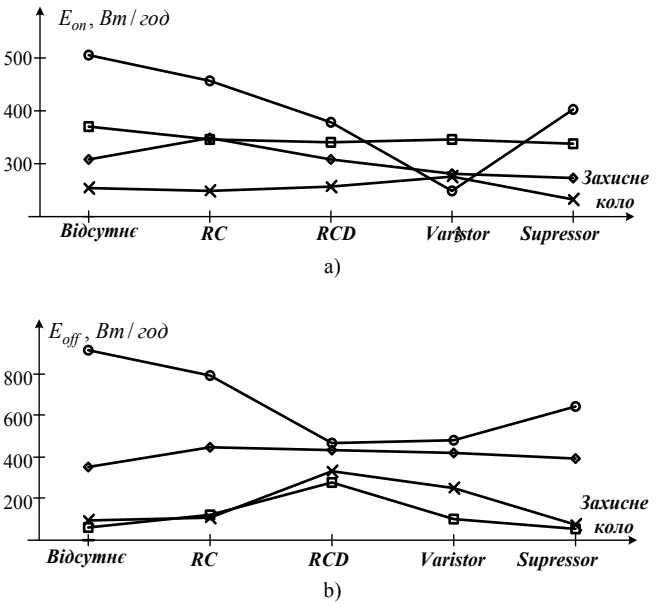


Fig. 9. Energy losses for IGBT turning-on (a) and turning-off (b) modes.

The characteristics of IGBT safety circuits usage for energy converter work with active-inductive load presented in fig. 10–12. In this case, same as it is for converter work with active load, overvoltage while IGBT turning-on doesn't

depend on safety circuit scheme and determines just parasitic inductance position.

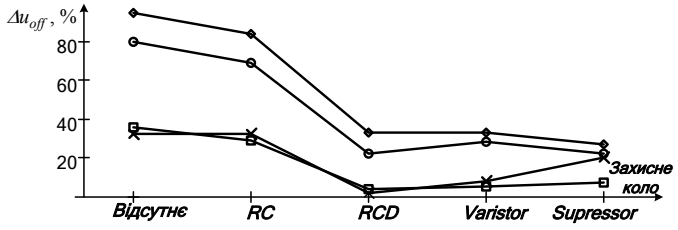


Fig. 10. Overvoltage dissipation in IGBT turning-off mode.

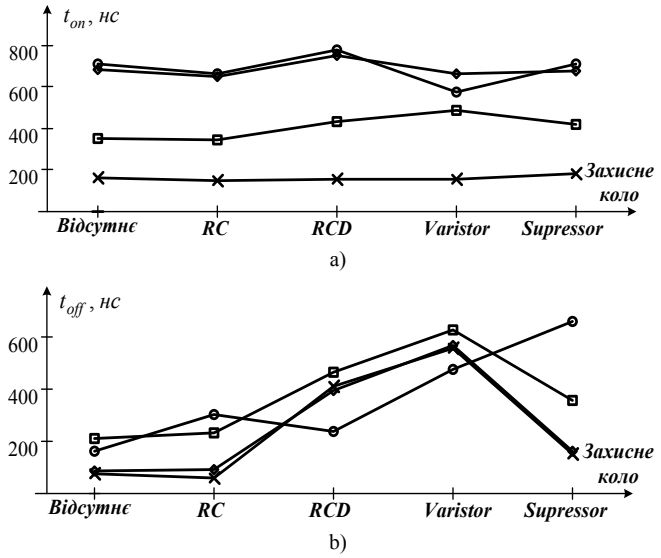


Fig. 11. Turning-on (a) and turning-off (b) IGBT time.

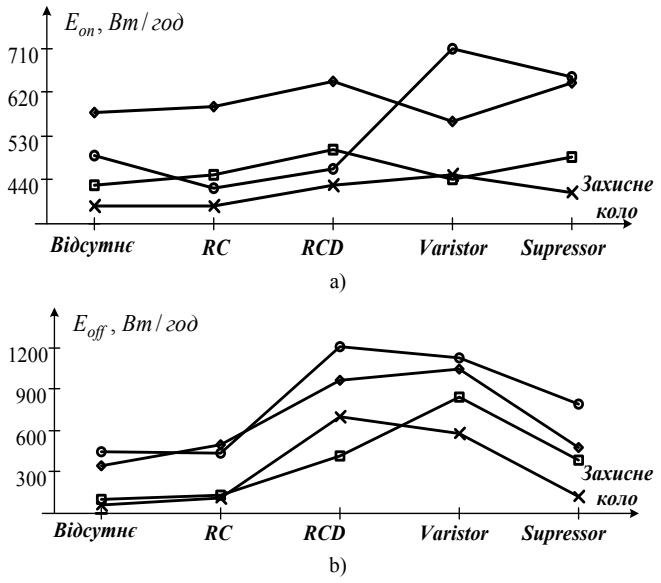


Fig. 12. Energy losses in IGBT turning-on (a) and turning-off (b) modes.

Thus, usage of additional transistor safety schemes in some case allows to solve task of overvoltage decreasing on a key. However, if safety circuit characteristics are not properly chosen, a significant deterioration of transistor operational modes appears. Thus, for not properly chosen characteristics of safety varistor the transistor couldn't be fully closed and there are voltage variations and current flow thru key in transients (fig. 13). In such mode energy losses increases significantly, so key heating is more intensive. This leads to IGBT fast breakdown.

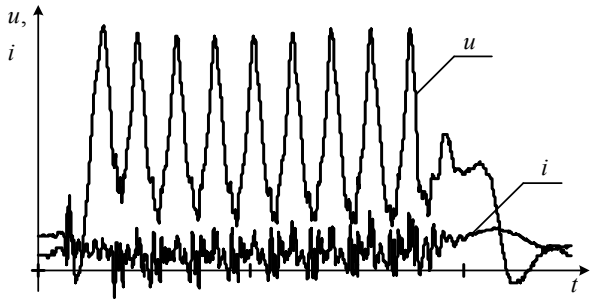


Fig. 13. IGBT current and voltage for not properly chosen characteristics of safety varistor.

VI. Conclusion

Usage of additional circuits of IGBT working point formation allows us provide split control of switching-on and switching-off speed of power keys, however it leads to increasing of energy losses on switching processes, and energy dynamic losses in transistors depends significantly on the characteristics of additional circuits elements. Usage of overvoltage safety circuits leads to significant deterioration such IGBT switching parameters as turning-on and turning-off time, which leads to decreasing energy losses in transistor switching modes. Evaluation of energy losses allows determining IGBT-based key heating for tasks of semiconductor power key designing.

References

[1] Воронин П. А. Силовые полупроводниковые ключи: семейства, характеристики, применение / П. А. Воронин. – [Изд. 2-е, перераб. и доп.]. – М. : Издательский дом Додэка–XXI, 2005. – 384 с.

[2] Мелешин В. И. Транзисторная преобразовательная техника / В. И. Мелешин. – Москва : Техносфера, 2006. – 632 с.

[3] Гейтенко Е. Н. Источники вторичного электропитания. Схемотехника и расчет : учебное пособие / Е. Н. Гейтенко. – М. : СОЛОН-ПРЕСС, 2008. – 448 с

[4] Семенев Б. Ю. Силовая электроника: от простого к сложному / Б. Ю. Семенев. – М. : СОЛОН-ПРЕСС, 2008. – 416 с.

FOC Control of PMSM Using Rapid Prototyping in MATLAB

[1]*Viktor ŠLAPÁK (1st year),* [2]*Michal PAJKOŠ (1st year)*
Supervisor: [3]*František ĎUROVSKÝ*

[1,2,3]Dept. of Electrotechnics and Mechatronics, FEI TU of Košice, Slovak Republic

[1]viktor.slapak@tuke.sk, [2]michal.pajkos@tuke.sk, [3]frantisek.durovsky@tuke.sk

***Abstract*—The paper present simple method of testing control structures on microcontroller device using rapid prototyping with MATLAB/Simulink. This approach is illustrated on example of PMSM field oriented speed control with speed encoder. Experimental results are presented and advantages and disadvantages of such an approach are described.**

***Keywords*—digital signal controller, PMSM, rapid prototyping**

I. Introduction

The permanent magnet synchronous motor (PMSM) is widely used in applications, which require high dynamic performance and accurate positioning. Its compact size predetermines PMSM to be used especially in robotics, but it can also be used in electrical vehicles as traction drive mounted directly to wheels. Therefore is current research focused on position control as well as speed control of PMSM. However, almost all of control structures are implemented in rotational coordinates. Except other control strategies, two high-performance control strategies, direct torque control and field oriented control are the most popular ones [1].

Development of such control usually starts with simulation of mathematical model of controlled plant in suitable simulation environment, e.g. MATLAB/Simulink. A complexity of mathematical model depends on type of control structure and available measuring sensors for given task. Therefore the implementation of designed structure could be difficult and however accurate and complex mathematical model is used, the implementation of designed structure could uncover facts, which will force change of parameters in the control structure or even change the structure itself.

As designed control is often implemented in microcontroller unit [2],[3], the need of transcription of control structure in programming code is obvious. For that purpose, programming language such as C, C++ or Assembler is used. Creating programming code is time consuming and resulting code is often very extensive and, therefore, disarranged. Creating programming code directly from simulation can be very helpful in case of quick experimental verification.

In this paper simple FOC control of PMSM using traditional PI controllers and its implementation is presented Control structure is designed in MATLAB/Simulink and almost same model is used to generate code for microcontroller unit.

II. Model of Permanent Magnet Synchronous Motor

Mathematical model of PMSM in rotor reference frame (dq-frame) is described according to [4], [5]. Electrical equations are:

$$\frac{di_d}{dt} = \frac{1}{L_d}u_d - \frac{R}{L_d}i_d + \frac{L_q}{L_d}p\omega_r i_q \tag{1}$$

$$\frac{di_q}{dt} = \frac{1}{L_q}u_q - \frac{R}{L_q}i_q - \frac{L_d}{L_q}p\omega_r i_d - \frac{\lambda p\omega_r i_q}{L_q} \tag{2}$$

$$T_e = 1.5p\left[\lambda i_q + \left(L_d - L_q\right)i_d i_q\right] \tag{3}$$

where: i_d, i_q – components of stator current in dq-frame, u_d, u_q – components of stator voltage in dq-frame, L_d, L_q – stator winding inductances in dq-frame, R – stator winding resistance, λ – flux amplitude inducted in stator, p – number of pole pairs, ω_r – angular velocity of the rotor, T_e – electrical torque.

Mechanical equations are:

$$\frac{d}{dt}\omega_r = \frac{1}{J}\left(T_e - F\omega_r - T_m\right) \tag{4}$$

$$\frac{d}{dt}\theta = \omega_r \tag{5}$$

where: J – moment of inertia, T_m – load torque, F – viscous friction, θ – rotor angular position.

For a symmetrical rotor is valid:

$$L_d = L_q = \frac{L_{ab}}{2} \tag{6}$$

where: L_{ab} – terminal inductance.

III. Design of FOC Speed Control

PMSM is controlled in closed speed loop with encoder. Stator current components are controlled by inner control loops which enable independent control of motor torque (created by q-component of current) and partially flux (created by d-component of current), if necessary.

Due to permanent magnets on the rotor, no flux created by i_d, is usually required. Therefore, the i_d component is kept on zero by PI controller. Negative value of i_d causes weakening of rotor flux, which is applied to achieve higher rotor angular speeds. Nevertheless, this method is not effective, because even a small change of speed requires a large stator current and enhanced heating of motor and therefore it is not covered in this paper.

Considering zero d-component of current, motor torque is generated only by i_q and flux linkage as seen in equation 3. Current i_q is controlled by PI controller, where its setpoint is set by speed PI controller and limited to maximum continuous motor current.

There are two compensations added to obtain decoupling of both current components:

$$u_d = u_{dcon} - L_q p \omega_r i_q \tag{7}$$

$$u_q = u_{qcon} + L_d p \omega_r i_d + u_{emfcomp} \tag{8}$$

$$u_{emfcomp} = \lambda p \omega_r \tag{9}$$

where: u_{dcon} – output of i_d current controller, u_{qcon} – output of i_q current controller, u_{emf} – EMF voltage, $u_{emfcomp}$ – compensation of EMF voltage

Based on the given equations, simulation model of controlled PMSM was created in MATLAB/Simulink environment. Simulated waveforms are presented on Fig.1 and simulation model is shown on Fig.2.

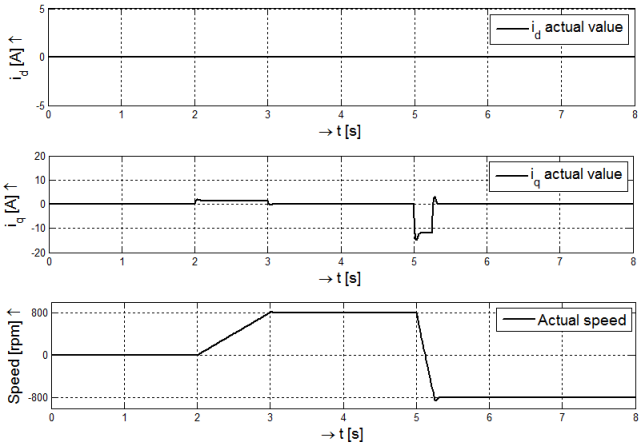


Fig. 1. Simulated waveforms of controlled PMSM

As can be seen in Fig. 1, both current components are controlled independently due to used compensations.

IV. Implementation

Control structure is applied on system with PMSM and a 3-phase inverter, which are controlled by digital signal controller (DSC). As was mentioned above, control structure was translated to programming code in C programming language from MATLAB/Simulink block model, which core was taken from simulation. However, few changes have to be made to adjust origin model for code generating.

First of all, specific blocks for DSC peripherals, as PWM modulation, AD converter and encoder evaluation need to be added to block scheme. On the other hand, PMSM model has to be removed from block scheme, because all signals are measured by sensors, therefore they do not need to be specially computed or simulated.

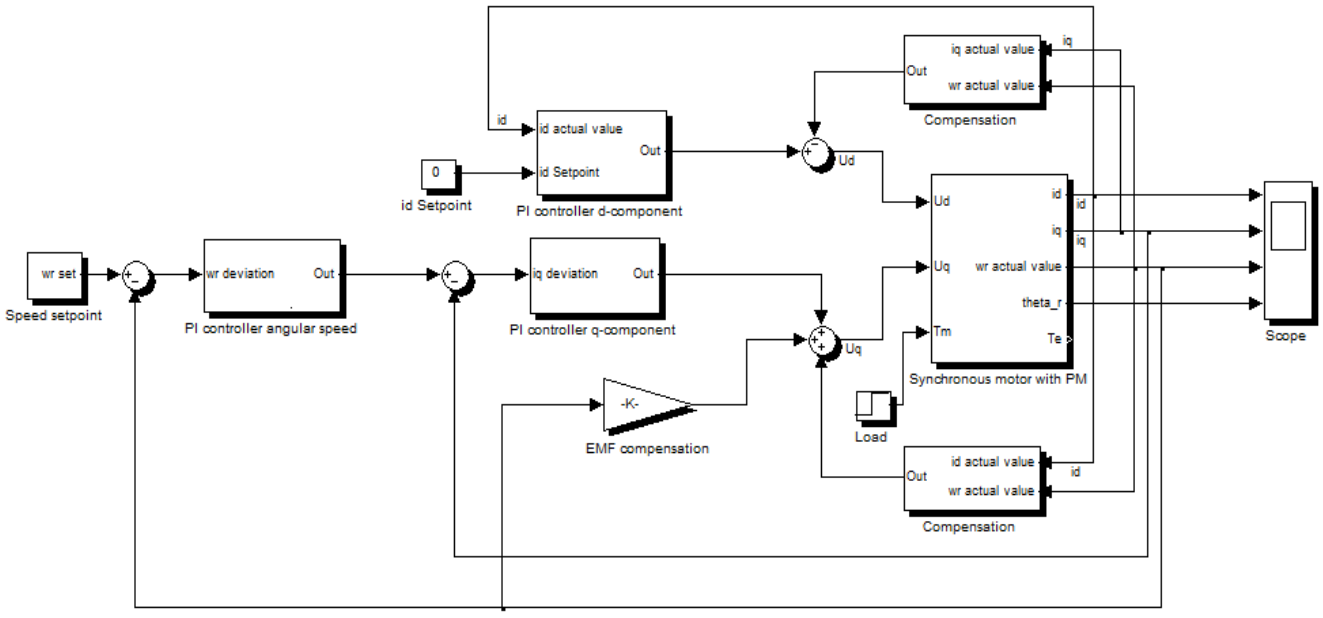


Fig. 2 Model of FOC control structure with PMSM

The most important block is *Target preferences* block, which sets main DSC parameters, simulation parameters to discrete with fixed sample time and also establish connection with DSC integrated development environment (IDE). In this implementation C2000 family DSC was used with Code Composer Studio v3.3 as IDE. After connection is established, MATLAB generates C code to IDE and it is also possible to give commands to DSC from MATLAB command line. Nevertheless, all of possibilities described are available only with addition toolboxes (e.g. Embedded coder and other toolboxes associated with it).

For peripheral control, appropriate blocks have to be chosen. Part of block scheme, PWM generation, is illustrated on Fig. 3. As can be seen, voltage commands are scaled and sent to PWM peripheral to be compared with its time base counter to create transistor switching signals. Figure 3 illustrates the easiness of peripheral control of DSC in Simulink.

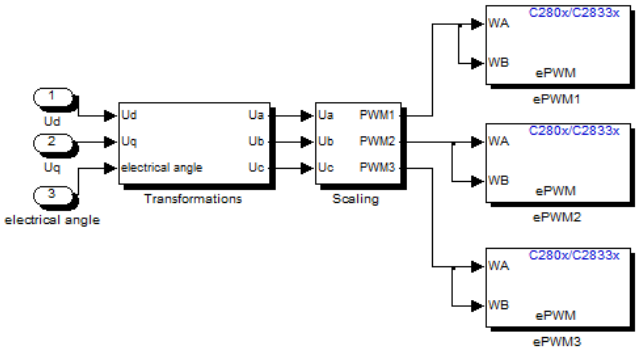


Fig. 3 Output voltage creation (PWM subsystem)

PWM module is represented by block, which is easy to set up without need of exact DSC register description, because Simulink block offers graphical block properties. Same approach is used to configure current sensing with current transducers connected to AD converter and also to configure quadrature encoder peripheral module used for position and speed sensing.

Necessary condition of code generating is that all blocks have to be discrete time blocks and correct sampling times must be set. In presented structure two sampling times were chosen. Current control loops together with position and current sensing sampling times are set to 0.5 ms whereas speed control loop, speed sensing and setpoint are sampled with 5 ms period. Speed control loop sampling time is a bit larger due to available speed encoder, which has low resolution. Smaller sampling time on the other hand would cause inaccuracy in speed evaluation.

Depending on DSC type, correct number format should be chosen. For fixed-point DSC using of fixed point number format is obvious and it is recommended to use special blocks from IQ Math library for multiplication, divide, frame transformation, etc. This approach enables less time consuming operation than with using standard blocks. In our application, floating-point DSC was used and therefore there was no need to use such blocks.

Using mentioned Simulink blocks the complete control structure is obtained and presented on Fig. 4.

Measured currents and speed were sent to MATLAB via RTDX (Real-Time Data Exchange) communication. It is unique type of communications, which change data with DSC in its idle time. It means that data exchange does not disturb control process. Disadvantage of this type of communication is that data are sent asynchronously with various sample times dependent on actual controller task. However, advantage of RTDX is its easy implementation. It consists from opening and closing communication channels and simple data splitting, if necessary.

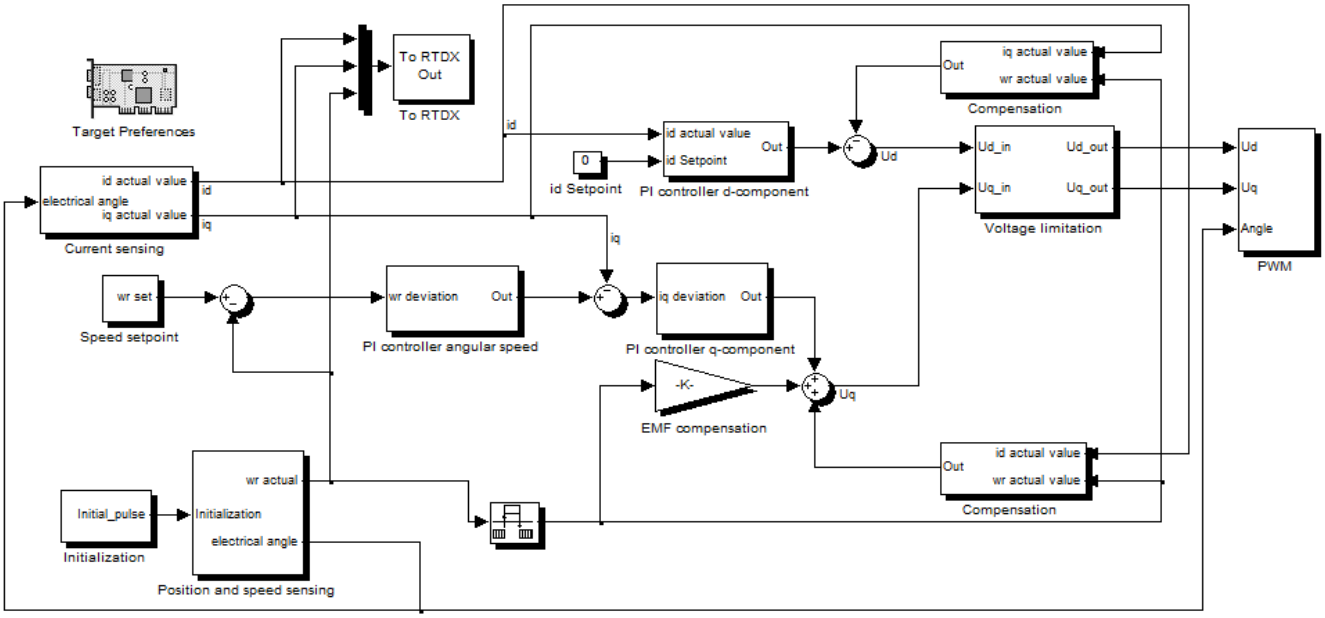


Fig. 4 Block scheme to be translated to C programming language

V. Experimental Results

Experimental results of motor control with no load, measured with DSC and sent to PC via RTDX, are presented in Fig.5. As can be seen in this figure, during motor running, controller maintains i_d current at zero whereas i_q current covers friction torque.

In 14-th sec. sharp reverse was performed, which cause negative peak of i_q current covering dynamic torque. Small peak of i_d during reversation shows inaccuracy of used compensations caused by difference between real and assumed motor parameters and also by inaccuracy of used speed encoder as they are used in compensation, equation (7).

Speed response corresponds to the simulated waveforms according to Fig. 1.

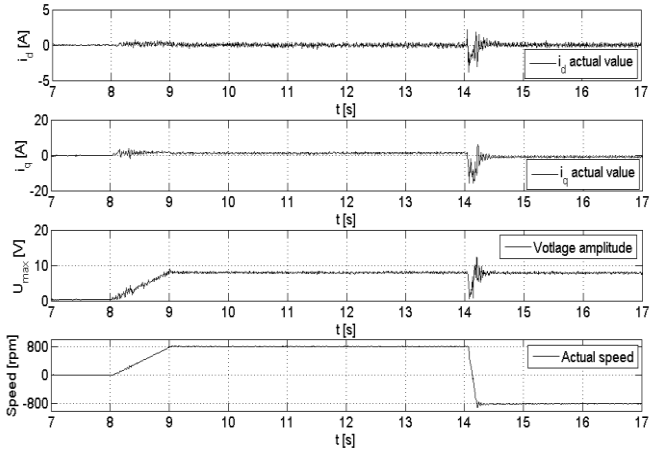


Fig. 5 Experimental waveforms

VI. Conclusion

Field oriented control of permanent magnet synchronous motor, designed and implemented as presented gives satisfactory results.

Great advantage of such an approach is the simplicity of implementation and mainly the possibility to quickly adjust control structure and following verification. It is also more user friendly programming than classical code writing, because it gives more natural perspective at designed structure.

Disadvantages include final C-code disarrangement and ineffectiveness caused by user ambition of transparent block structure and therefore, for example, using multiplication where it is not necessary.

To sum up, rapid prototyping using MATLAB/Simulink is useful in fast development or verification and also to help students with practical implementation of their work.

Acknowledgement

This work was supported by the Slovak Research and Development Agency under the contract No. APVV-0185-10.

References

[1] M. S. Merzoug, F. Naceri," Comparison of Field-Oriented Control and Direct Torque Control for Permanent Magnet Synchronous Motor (PMSM)", World Academy of Science, Engineering and Technology 45 2008

[2] Bilal Akin, Manish Bhardwaj, Jon Warriner, "Sensorless Field Oriented Control of 3-Phase Permanent Magnet Synchronous Motors", Texas Instruments, Mar 2011

[3] Milan Brejl," Permanent Magnet Synchronous Motor with Resolver, Vector Control, Driven by eTPU on MPC5500", Freescale Semiconductor, 03/2012

[4] Grenier, D., L.-A. Dessaint, O. Akhrif, Y. Bonnassieux, and B. LePioufle. "Experimental Nonlinear Torque Control of a Permanent Magnet Synchronous Motor Using Saliency." IEEE Transactions on Industrial Electronics, Vol. 44, No. 5, October 1997, pp.680-687.

[5] R. Krishnan,"Permanent Magnet Synchronous and Brushless DC Motor Drives",Virginia, U.S.A, CRC Press, 2010

Impact of the S – parameters to the Printed circuit board characteristic in the UWB application

[1]*Ing. Matej ŽIGA (1st yearPhD.),* [2]*Bc. Martin JANUŠ (2nd year Ing.)*
Supervisor: [3]*Pavol GALAJDA*

[1, 2, 3]Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

[1]matej.ziga@student.tuke.sk, [2]martin.janus@student.tuke.sk, [3]pavol.galajda@tuke.sk

***Abstract*— The well-known advantages of Silicon – Germanium bipolar technology, like low cost, small size, or high integrability, have recently been expanded to highest- frequency applications.[9] The article presents in the article is presents M – sequence based ultra – wideband (UWB) sensing system for the Federal Communications Commission and Electronic Communications Committee frequency bands Our design consists from the Printed circuit board and ASIC chip. The packaged ASIC chip consists of down – conversion active mixers. The mixers are based on conventional double-balanced Gilbert topology. For its characterization, the chip has been mounted on a test board to RO4003C which were observed Scathering parameters.**

Keywords—BiCMOS, UWB, S – parameters, IE3D, MGRID, mixer

I. Introduction

UWB technology has several applications in different areas necessary for our lives. The main feature this technology is use of wireless communication to transmit information. The classical task of information transfer at high frequencies in the GHz can browse through different materials, which can cause interference with other services, e.g. TV – and audio broadcasting, mobile communication (GSM, wireless networks, etc.), and GPS.[10] In modernizing a high frequency technologies needs to improved several parameters. The most significant observed parameters are the Scathering parameters (S – parameters), which have an important role in the design of printed circuit board (PCB).

S - parameters forms the matrix consisting of 4 elements (S11, S12, S21 and S22). With these elements can be observed crosstalks between signals and impedance matching of the conductive paths. Between another the significant parameters observed may be include gain, return loss and many others.

When the design PCB boards with S - parameters to be taken into considerations a frequency band for the UWB applications. A broadband ultra wideband (UWB) technologies uses in the United States according Federal Communications Commission (FCC) regulations a frequency band from 3.1 GHz to 10.6 GHz and in the Europe frequency band from 6 GHz to 8.5 GHz released institution of the Electronic Communications Committee (ECC).[7][8]

One of the most growing solutions of the UWB radar systems is M - sequence. The UWB radar system consists of several parts which can be placed on the printed circuit board (PCB). The basic parts of the UWB M - sequence radar are the M - status shift register, mixer, I - Q demodulator, analog - digital converter (A / D) and processor. The block diagram of the UWB M - sequence radar is on the Fig. 1

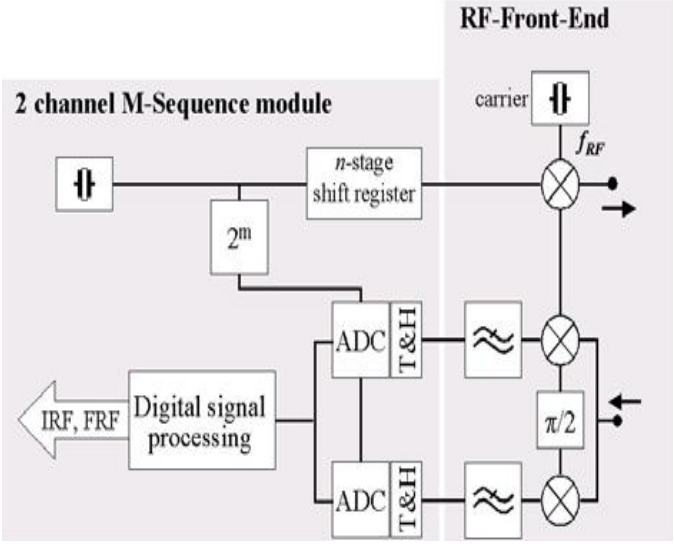


Fig. 1 Block diagram M – sequence UWB radar[4]

One of the first blocks of signal processing in the UWB M sequence radar is a mixer, which was designed at the Technical University in Kosice for band the ECC. For correctly definition parameters of a mixer is necessary to matching the PCB for all the S - parameters. The aim this paper is clarify the reader how have affected S-parameters of the PCB where is necessary to provide nearly ideal conditions for the transmission of high-frequency signal.

The paper is divided into three parts. The first part deals with the theory of S-parameters. The second part describes the PCB design for mixer. Last section shows the implementation of the proposed PCB.

II. Basic theory of S - parameter

S - parameters can be characterized by two port network. Two-port network are characterized by two input pins and two output pins.

The measurement of S - parameters means to send a small signal into two - input equipment and the answer of input, and output pins is recorded and described in graphic dependence.

S - parameters can be characterized by two port network. Two-port network is characterized by two input pins and two output pins. Measuring S - parameters for two-port is to send a small signal at the input and to the output pin monitoring of their change. These parameters are defined in the therms of travelling waves and completely characterize the behavior of two-port networks. Through the use of these variables, we can describe a family of parameters Z, Y, H, and S parameters. S - parameters represent a linear combination of voltages and currents which are dependent and independent variables.[2]

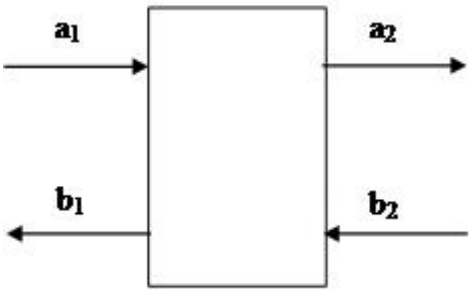


Fig. 2 Description two-port using auxiliary parameters

$$a_1 = \frac{1}{2\sqrt{z_0}}\left(v_1 + z_0 i_1\right)\ [2] \quad (1)$$

$$a_2 = \frac{1}{2\sqrt{z_0}}\left(v_2 + z_0 i_2\right)\ [2] \quad (2)$$

$$b_1 = \frac{1}{2\sqrt{z_0}}\left(v_1 - z_0 i_1\right)\ [2] \quad (3)$$

$$b_2 = \frac{1}{2\sqrt{z_0}}\left(v_2 - z_0 i_2\right)\ [2] \quad (4)$$

In the real connection backscattering coefficient measurement, in other words transient characteristics (crosstalk) is based on an alternating signal sent is mostly sinusoidal signal, which can change the size of the amplitude. This signal is sent from the input to the output parameter S12, where the signal is recorded and plotted on the output pin. The opposite case is rendered parameter S21.

III. PCB Design

When designing the PCB we assume theoretical knowledge of the impact of S - parameters which must also take into account the size and layout of input - output pins. PCB board shown in Fig. 3 is composed of ASIC mixer wired with input and output matched connectors.

A. Mixer

For the purposes of M - sequence UWB radar was designed monolithic chip which containing a pair of mixers.

The mixers are based on the double-balanced Gilbert cell. Consumption of the whole chip is 75mA and is connected at the negative supply voltage -3.3 V.

Chip with mixer is designed and manufactured in 0.35 um SiGe BiCMOS technology, where size of silicon die is 850 x 1950 um. Chip bonding is mounted in the housing 32 on the QFN 5 x 5 mm.

B. Simulation in the application HyperLynx® 3D EM

Our main goal is use electromagnetic simulation, which provides a very accurate analysis of complicated high-frequency designs, antennas, high-speed circuits. Before simulation of the simulation program is necessary to design the first version of mixer PCB board in design the Altium Designer application. This is a complex and redesigned environment where is possible realize easily the design of electronic circuit drawing after PCB design. The proposed board is converted into simulation program the Hyper Lynx 3D EM from the Mentor Graphics. The IE3D program package allows to create complicated structures, vias and grounding plates of different thicknesses. At the importing into simulation program the IE3D was used the MGRID program (Fig. 5), in which we can create and modify the structure. This is divided on polygons and points. The program includes a simulation through which we can view the S - parameters of generated structures.

In this program, we simulated the structure proposed by us. After the simulation of results, was modified structure of individual lengths of conductive lines. After further simulation we added further vias between a individual conductive paths.

On the basis of these changes was created the second version of the mixer. In this version was reduced length of conductive paths and at the whole PCB board was increased number of vias.

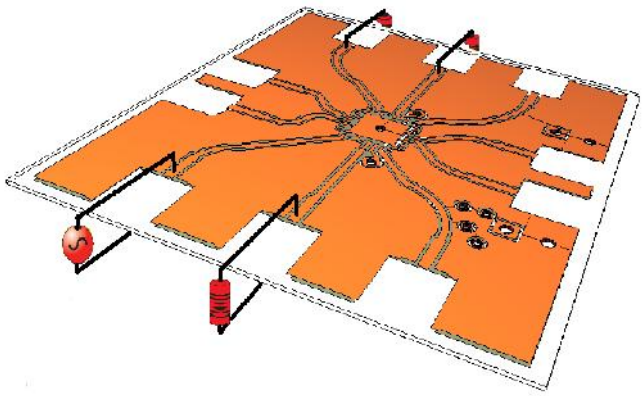


Fig. 3 The principle of simulation in IE3D CAD tool

C. Results of the simulation from IE3D

At the simulation in simulation program IE3D, it is first necessary to define the port simulation. In this program is can choose from seven types of ports. The program MGRID we defined a port called Advanced Extension. Its typical value of impedance is 50 Ω ports, which are used at adapt of impedance the transmission of leadership.

At the testing the first version of mixer, we simulated the transmission paths between the ports called IF-LO. Because of the robustness of DPS design was necessary to carve a design. The simulation was performed with step 0.16 Hz. After simulation was made this board. Comparison of the results between simulated and measured results are shown in the Fig. 4. The measured results show that a measured results achieved better values than results of simulation. This difference could be caused by not accepting all the vias transmission path. At

the simulation of the second version of mixer was this problem removed at the simulation. The length of simulation was 7 times longer than a simulation in the first version of mixer. The result of this simulation are results comparable with the measured results in the first version of mixer.

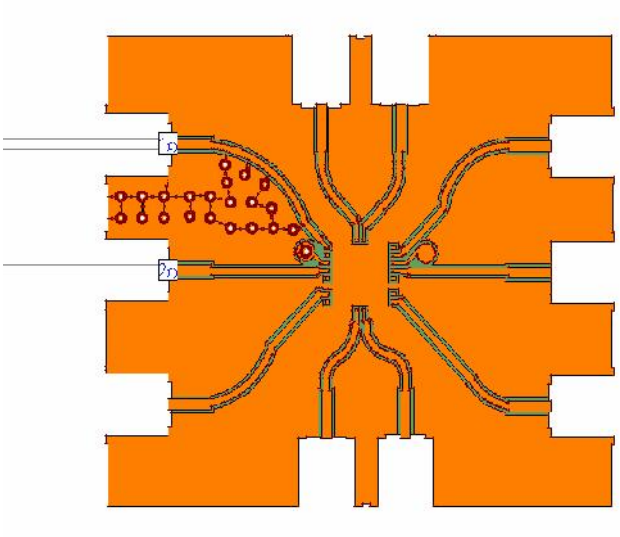


Fig. 4 Simulation in the MGRID

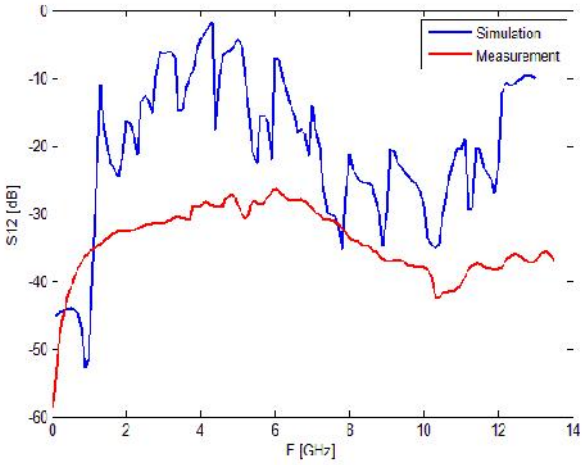


Fig. 5 The comparison simulation and measurements from the first version of the PCB

IV. Implementation of the Design

The first version PCB board was made at the substrate RO4003C double layer PCB board. Based on simulations of the creation results will be transformed in the new version PCB board. New version will be implemented at the RO4003C, where relative permittivity is 3,55 and thickness is 0,203 mm.

V. Conclusion

The aim of the study was effort to underline the impact of crosstalk high frequency PCB production, which is used in the UWB radar system. In a future version of this board is necessary to reduce crosstalk between single conductive paths.

References

[1] LIPTAJ, Martin, ŽIGA, Matej, GALAJDA, Pavol: Design of the IQ-modulator for UWB applications. Košice - FEI TU, SCYR 2012, 4 p.
[2] S – Parameters Circuits Analysis and Design, Application note 95, Semptember 1968.
[3] ROBENS, Markus, WUNDERLICH, Ralf, HEINER, Stefan : UWB LNAs for Ground Penetrating Radar. IEEE Avaible on internet: 12.4.2012 http://ieeexplore.ieee.org/ieee_pilot/articles/5076158/05117727/article.html#authors
[4] Website subject UWB bezdrôtové senzorové siete: https://moodle.tuke.sk/moodle202/mod/resource/view.php?id=1661
[5] Datasheet RO4003C od firmy Rogers. Avaible on internet: 27.4.2012 http://www.rogerscorp.com/documents/726/acm/RO4000-Laminates---Data-sheet.aspx
[6] User manual HyperLynx® 3D EM IE3D, 2012, 658 p.
[7] FCC News Release :New Public Safety Applications and Broadband Internet Access Among Uses Envisioned by FCC Authorization of Ultra – Wideband Technology
[8] Electronic Communication Committee: ECC Decision of 24 March 2006 amended 6 July 2007 at Constatnta on the harmonized conditions for device using Ultra Wideband (UWB) technology in bands below 10,6 GHz
[9] LIPTAJ, Martin, KMEC, Martin, GALAJDA, Pavol: A SiGe Based Extension Kit for Ultra Wideband ECC Compliant Sensing System, 6 p.
[10] HERMANN, R., SACHS, J., KMEC, M., GRIMM, M.: Ultra – Wideband sensor system for remote monitoring of vitality at home, 4p.
[11] GONZALEZ, Guillermo : Microwave transistor amplifiers: analysis and design, Prentice – HALL, Inc.,1997, 516 p., ISBN 0-13-254335-4

Improvements in reliability of through-wall detection of static persons by UWB radar

[1]*Peter KAŽIMÍR (1^{st} year),* [2]*Jana ROVŇÁKOVÁ*
Supervisor: [3]*Dušan KOCUR*

[1,2,3]Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

[1]peter.kazimir@tuke.sk, [2]jana.rovnakova@tuke.sk

***Abstract*—This paper describes a signal processing procedure for through-wall detection and localisation of static persons by ultra-wideband (UWB) sensor and discusses selected factors influencing the reliability of this technique. UWB sensor described in this paper consist of the M-sequence sensor, one transmitting and two or four recieving antennas, data acquisition and data processing software tools. Experimental measurements with static person in three different positions proved the significance of selected factors. Improvements in reliability of target detection are discussed in this paper.**

***Keywords*—Breathing, detection, localisation, spectral density, through-wall, ultra-wideband.**

I. Introduction

The UWB sensors are becoming a wide-spread trend in sensorics. Electromagnetic waves (even those of very low power) emitted by UWB sensors have good penetration through most common building materials but are reflected from human body. UWB sensors also offer high spatial resolution. This attribute allows several approaches to use of UWB sensors to be defined – ranging from military, security and rescue purposes through medicine to commercial use. Some of the recent research topics include through-wall detection of trapped people [1] [2], sensing heartbeat and breathing rate or detection of moving targets behind the obstacles [3],[4].

Detection of moving persons, as well as detection of static persons is based on the observation of the target area by stationary UWB radar system and extracting the information about how the acquired impulse responses changed over time. While detection of moving targets is based on the detection of non-stationary signal components in the time domain, the detection of static persons is based on the detection of respiratory motions in the frequency domain [5].

In the detection of static persons whose limbs are motionless, we can still observe geometric alterations of body shape caused by inner organs (lung, heart). These changes of body shape are discernible by a high resolution UWB radar. In positioning of static persons a target detection is very challenging by itself. It is based on the periodical nature of breathing or hearth beat that makes it possible to distinguish it from noise and clutter components.

This paper has following organisation. Section II contains M-sequence sensor [6] technical overview. Section III describes the signal processing procedure for detection and localisation of static persons based on the detection of human breathing rate. Section IV contains a discussion on selected factors influencing the reliability and usefulness of this technique. Results of experimental measurements for three different scenarios are given in Section V and conclusion is presented in Section VI.

II. M-sequence UWB sensor

The UWB M-sequence sensor is a prototype sensor originating from the Ilmenau Laboratory for Microwave Measurements and Sensors and was developed by ILMSENS team. M-sequence sensor is very resistant to narrowband interference. The sensor used in this paper is a 9th order M-sequence sensor which can operate in range up to 17 metres with 3cm spatial resolution. This sensor includes an UWB signal generator which generates wideband pseudo-random codes, and two synchronously operating sub-sampling receivers. One transmitting and two recieving antennas are required for proper operation of the radar system because the trilateration method is used in the signal processing procedure. The sensor communicates with the computer via Ethernet connection. The whole system setup used in experimental measurements described in this paper is shown in the Fig. 1.

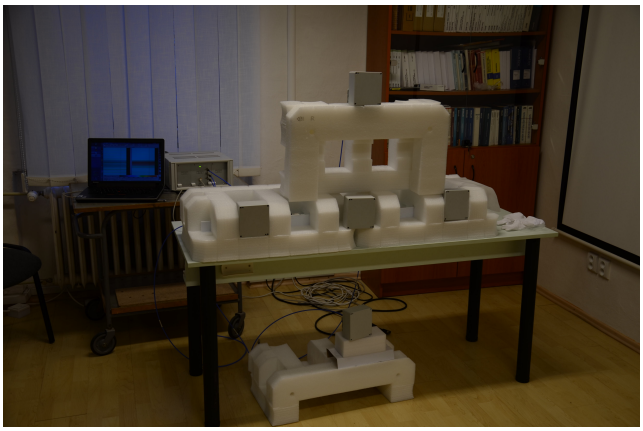


Fig. 1. UWB sensor tabletop unit in experimental measurement setup.

III. UWB sensor signal processing

The base of signal processing procedure for static persons positioning was originally introduced in [7]. It is similar to the procedure of detection of moving persons except that body movements are markedly restricted. Therefore, the signal to clutter ratio will further decrease and the only feature to distinguish the persons from static objects is their small

movement due to breathing. There are few features that serve as a basis for the methods of enhancing the response from breathing person:

- Periodicity of human breathing - the frequency of breathing is usually between 0.2-0.5 Hz and changes slowly over time;
- The geometrical variations of the thorax caused by breathing are usually quite less than the range resolution of the radar;
- The echo due to breathing person is extremely weak, and is even more attenuated when the static person is behind obstacle (e.g. rubble, wall, snow);
- Position of the antennas does not change during the measurement.

Taking into account these features, the signal processing procedure for detection of static persons starts with the phase of breathing enhancement and is followed by transformation of the radar signals into the frequency domain. Consequently, the estimation of power spectral density is used for breathing detection task. The remaining processing phases are time of arrival (TOA) estimation, wall effect compensation and target localisation and are the same as in the procedure for positioning of moving persons. The original (raw) data acquired from the sensor are often displayed as radargram, example of this radargram is shown in Fig. 2.

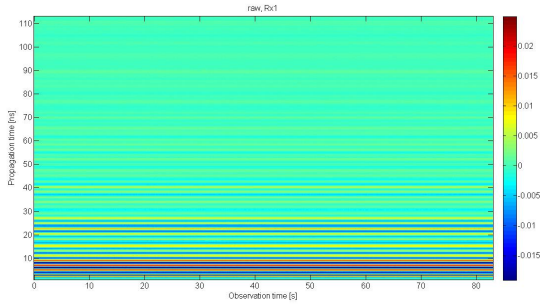


Fig. 2. Radargram of raw radar data.

A. Background subtraction and breathing enhancement

The method that subtracts the stationary background and clutter from the signal is reffered to as exponential averaging method. It is used for both moving and static persons. However, the weighting factor controlling the amount of averaging in the background estimation should be set to smooth out high frequency variations and reveal long term trends in the background estimation (i.e. it provides low-pass narrow band filtering). It is done by choosing a longer fraction of the previous estimate of impulse response with subtracted background and a smaller fraction of the actual measured impulse response. In order to further improve the signal-to-noise ratio of a static target echo, we apply range fileter on the impulse response with subtracted background before the target detection. The range filter helps to improve the SNR by reducing the clutter residue and noise resulting from the decorrelation of any radio frequency interference due to pseudo-random code transmitted by the radar. Additionally, breathing as a narrow band process can be enhanced by low-pass filtering utilization. Here, a low-pass filtering with cut-off frequency higher than the highest frequency of breathing (e.g. higher than 1 Hz) can be applied along the observation time axis for each propagation time instant to suppress high-frequency noise. Examples of background subtraction and breathing enhancement results are shown in Fig. 3.

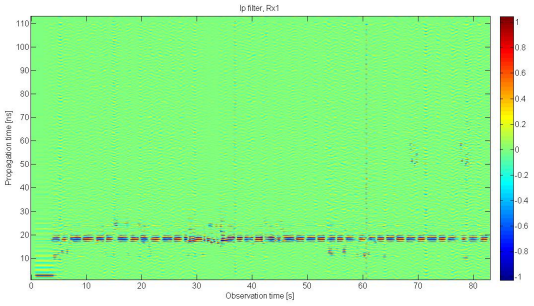


Fig. 3. Radargram of data after background subtraction and filtering.

B. Estimation of power spectral density

To extract the breathing rate a horizontal Fast Fourier Transform (FFT, along the time observation axis for each propagation time instant) is applied on the radar signals after background subtraction and breathing enhancement. The frequency of breathing can change with the observation time. However, the bandwidth of breathing from one person under observation with radar is likely to be considerably less, then a priori bandwidth as determined for the whole range of respiratory activity for all individuals. Thus, the total energy contained within the frequency window can serve as an indicator whether breathing is present. Finally, FFT-based Welch periodogram is utilized for estimating the power spectral density (PSD) of the radar signals in the direction of the observation time. Examples of PSD estimation phase is shown in Fig. 4.

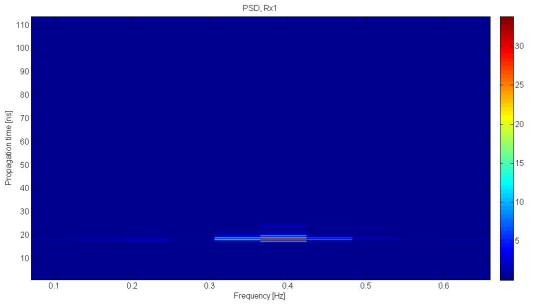


Fig. 4. Power spectral density of target breathing at 0.4 Hz rate.

C. Target detection

For detection of static persons, the constant false alarm rate (CFAR) detector or simpler threshold detector can be used. The detector is applied on data set represented by the estimated PSD. If a breathing person is present in the monitored area, the detector binary output should gain values 1 between frequencies 0.2-0.5 Hz corresponding with the expected breathing rate of human being as shown in Fig. 5. If more static persons are situated inside the area, values 1 should occur in more propagation time instants. In order to later estimate for every detected target only one spatial position, the frequency responses from interval 0.2-0.5 Hz are simply summed to one frequency response. Such response represents then input to the TOA estimation phase.

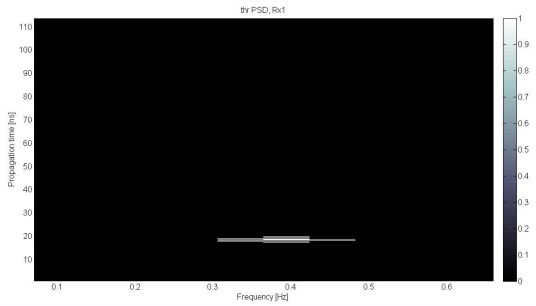


Fig. 5. Target detection with target breathing at 0.4 Hz rate.

D. TOA estimation

The TOA estimation phase is realized by TOA association method. Estimated TOA for every detected static target represents a round trip time between transmitting antenna target receiving antenna. TOA multiplied by the light propagation velocity gives the distance between them.

E. Target localisation

If both receiving antennas detect the presence of breathing person, its position will be calculated as intersection of ellipses given by the couple of TOA associated during TOA estimation phase. If only one receiving antenna confirms target presence, at least incomplete positioning of static person based on the distance from this antenna can be utilized. The possible locations of the person are then given by the half-ellipse situated inside a monitored area. Fig. 5 shows a position of successfully detected and located target.

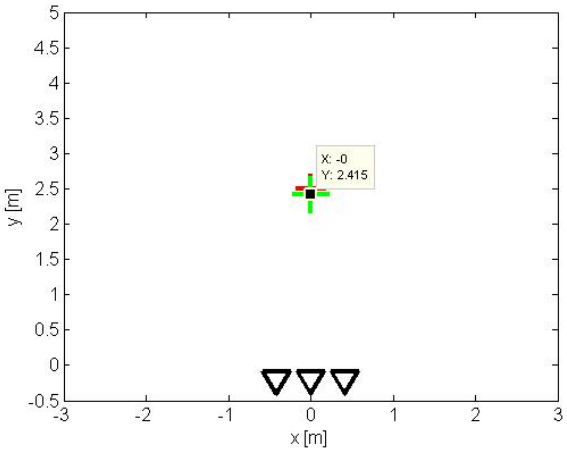


Fig. 6. Display of calculated position of static target (green cross) and real target position (red cross).

IV. Selected factors influencing detection of static person

This section describes selected factors influencing the reliability and precision of static human detection and localisation. When creating a sensor system for emergency and rescue purposes, we have to take into consideration that it should be easy to operate even by people without background on signal processing procedure. With regards to this fact, we chose these factors that can improve the system reliability: detector threshold level, measurement duration and positions of antennas. The impact of these factors on the static person detection was proved by experiments described in Section V.

A. Detector threshold level

When attempting to determine presence of the static person through the wall, we assume that we have no previous information about the target presence, position and location. We have to rely on the information acquired from the sensor and the results of signal processing procedure. The detector threshold level indicates whether the PSD of the acquired signal for each recieving antenna is high enough to consider the presence of the target. However, in different scenarios and different target body positions, the PSD of recieved signal changes significantly. Using very high threshold level may cause inability of sensor to detect a person even if it is present, on the other hand, very low threshold may cause false alarms and erroneous detection of persons.

B. Measurement duration

The periodical nature of breathing allows us to detect a static person position by signal analysis in spectral domain. The frequency of human breathing is relatively low and we need to acquire a data set with duration that varies, depending on the scenario, from several breathing cycles to few minutes for successful PSD analysis. If the system is aimed for use in emergency cases, it should be capable of detection of targets within the least possible time.

C. Position of recieving antennas

The recieving antennas are usually positioned in line with transmitting antenna and they determine the X-axis of 2D coordinate system, with transmitting antenna assigned to [0,0] position. However, this positioning is not universal and different body positions may cause different gain in PSD which may lead to unability to detect the target until the recieving antennas position is changed. Searching for the right positioning of antennas is time consuming and may cause the operator to lose focus on the task at hand.

V. Preliminary experimental examination of selected factors impact on system reliability

In this section we provide preliminary results of experimental measurements in three different body position scenarios and two different recieving antennas setups. In these measurements the person was located in the 2,5 m distance from the sensor in the line of sight. In the first scenario the person was sitting motionless, facing towards the transmitting antenna, this scenario will be reffered to as “direct”. In the second scenario the person was sitting motionless, rotated by 90°, this scenario will be reffered to as “side”. In the third scenario the person lays motionless, rotated by 90°, this scenario will be reffered to as “laying person”. Two antenna positions were examined, the first setup was commonly used horizontal setup with 0.42m distance between transmitting and recieving antennas, the second setup was vertical with 0.49m distance between transmitting and recieving antennas. Experiment was performed in the complex enviroment of laboratory equipped with lots of strong reflectors such as whiteboard, heater, electronic measurement devices.

A. Antenna positioning

Experimental measurements prove that antenna positioning has a significant influence on the reliability of target detection. We assume that target body position is unknown, we therefore have to guess the optimal antenna positioning. Two antenna setups were tested. The first, most common setup with antennas placed in horizontal line provided good results for "direct" target position, but the target presence was almost indefinable for "side" and "laying person" scenarios. On the other hand, the vertical setup, with recieving antennas placed above and under the transmitting antenna, provided good results in all of these scenarios, which means that the target presence could be detected. We propose a more complex but seemingly more reliable setup with use of all four antennas for the real-world application, as it improves the reliability of the system and solves the problematic issue of finding an optimal position.

B. Threshold level impact examination

As mentioned in Section IV, the PSD may change significantly according to the measurement scenario. If the detector threshold will be set very low, the detector will lose its purpose and unknown number erroneously detected targets might appear and cause confusion. The task of setting the detector threshold level to a reasonable value is even more complicated if there is inequality in PSD gain from the recieving channels. Such inequalities may occur for example if one of the antennas is closer to a large metalic object. This inequality can cause unability to detect the target presence, if both channels are analysed according to the same detector threshold level. For example, in the horizontal antenna setup for "side" scenario, the PSD peak for one recieving channel is approx. 3 times bigger than the second and the optimal threshold level for the channel one does not detect a person in channel two, and optimal threshold level for channel two causes loss of useful data in the channel one as shown in Fig. 7. We therefore present an adaptive detector threshold which is calculated separately from the maximum peak in PSD results for each recieving channel. According to the preliminary experimental measurements, we assume to set the addaptive threshold at 60% of maximum of PSD function.

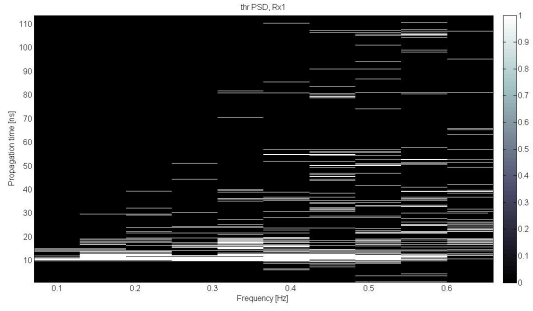


Fig. 7. Very low threshold level applied to PSD results.

C. Measurement duration impact examination

In the emergency situations, the minimum required measurement duration is very important factor. In the "direct" scenario, the breathing frequency was easily detected for both antenna setups. We were capable of detection and localisation of the target within 10s of the measurement, which corresponds to 135 impulse responses. This could be done for both fixed detector threshold and adaptive threshold. In the more complex scenarios "side" and "laying person" longer measuremet was needed. For "laying person" scenario with vertical antenna setup and fixed detector threshold it was neccessary to collect the data for over 60 seconds and it was first needed to find the optimal threshold manually. If the adaptive threshold was applied on this measurement, we were able to detect the target presence within 40 seconds of measurement. In the "side" we were able to detect the static person presence within 10 seconds of measurement with use of adaptive threshold, we were also able to detect the static person with use of fixed detector threshold, but manual search for the optimal detector threshold was required. In some cases we are able to determine the static target presence within seconds of measurement, but some scenarios require prolonged data acquisition for reliable detection. We propose the use of detector with adaptive threshold on the first 10-20 seconds of the measurement, and only perform longer measurements if the target presence is not confirmed. Results of preliminary experimental measurements in three different line of sight scenarios and two different antenna positions are given.

VI. Conclusion

This paper discusses selected factors influencing the reliability of static person detection via UWB sensor system. Modifications improving the usefulnes and reliability of the static person detection are proposed and proved by the preliminary results of experimental measurements. More specific signal processing procedure parameters that have impact on the static person detection are not in the scope of this paper.

VII. Acknowledgement

This work was supported by the Slovak Cultural and Educational Grant Agency (KEGA) under the contract No. 010TUKE-4/2012 and by the Scientific Grant Agency of the Ministry of Education, science, research and sport of the Slovak Republic and the Slovak Academy of Sciences (VEGA) under the contract No. 1/0563/13.

References

[1] E. Zaikov, J. Sachs, M. Aftanas, and J. Rovnakova, "Detection of trapped people by uwb radar," *Microwave Conference (GeMIC), 2008 German*, pp. 1 –4, march 2008.

[2] A. Nezirovic, A. Yarovoy, and L. Ligthart, "Signal processing for improved detection of trapped victims using uwb radar," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 48, no. 4, pp. 2005–2014, April.

[3] M. Chia, S. Leong, C. Sim, and K. Chan, "Through-wall uwb radar operating within fcc's mask for sensing heart beat and breathing rate," in *Microwave Conference, 2005 European*, vol. 3, oct. 2005, p. 4 pp.

[4] R. Zetik, J. Sachs, and R. Thoma, "Uwb short-range radar sensing - the architecture of a baseband, pseudo-noise uwb radar sensor," *Instrumentation Measurement Magazine, IEEE*, vol. 10, no. 2, pp. 39 –45, april 2007.

[5] J. Rovňáková and D. Kocur, "Uwb radar signal processing for positioning of persons changing their motion activity," *Acta Polytechnica Hungarica,*, 2013, accepted for publication.

[6] J. Sachs, R. Herrmann, M. Kmec, M. Helbig, and K. Schilling, "Recent advances and applications of m-sequence based ultra-wideband sensors," in *Ultra-Wideband, 2007. ICUWB 2007. IEEE International Conference on*, Sept., pp. 50–55.

[7] J. Sachs, E. Zaikov, D. Kocur, J. Rovňáková, and M. Švecová, "Ultra wideband radio application for localisation of hidden people and detection of unauthorised objects," D13 Midterm report on person detection and localisation, Project RADIOTECT, Tech. Rep., 2008.

Influence of short-time thermal degradation of vn stator coil to partial dicharge activity

[1]*Marián HRINKO (3^{st} year),* [2]*Lukáš LISOŇ (1^{nd} year)*
Supervisor: [3]*Iraida KOLCUNOVÁ*

[1,2,3]Dept. of Electrical Power Engineering, FEI TU of Košice, Slovak Republic

[1]marian.hrinko@tuke.sk, [2]lukas.lison@tuke.sk

***Abstract*—Insulation system is the most important electrical part of high voltage rotating machines. Its good condition contributes to reliable operation of machines. But during the time insulation system is subjected to aging. It causes that insulation become weaker and less resistant to possible stresses. Diagnostic is therefore very important step. The partial discharge measurement is one of the very important non-destructive diagnostic method to acquire information about quality of insulation system. This article describes partial discharge measurement and monitors how the change phase resolved partial discharge patterns of insulation system after the short-term thermal degradation of high voltage coil.**

***Keywords*—Partial discharges, thermal stress, high-voltage machines.**

I. Introduction

Insulation system of high voltage rotating machines during operation is exposed by many of different stresses that can affect the rate of insulation degradation in stator windings. These stresses are thermal, electrical, mechanical and ambient. Its act on insulation system differently but all of them after the given time decrease life of the insulation (aging). The aging causes a modification of the physical and chemical properties and the insulation became weaker. It can lead to electrical breakdowm of insulation system and this is the last state leading to failure of hole equipment. The monitoring is therefore important step in holding the insulation system in good condition. Reveal exceeded values any of these stresses acting on insulation system betimes should restrain faster aging and avoid the possible failures. This article describes electrical stress - partial discharges with the combination of thermal stress [1].

Partial dischrage can be characterized as small electrical sparks that don't bridge the distance between two electrode which originates on the surface or cavity where is enhanced field strength. If the cavity is filled with an organic solid or liquid, the partial discharges will degrade the organic material. Not inhibited of partial discherges can lead to reapeted sparks and this could eventually erode the hole insulation [2].

Thermal aging depends on the nature of the insulation (thermoset or thermoplastic) and the operating environment (air or hydrogen). In thermoset insolation, thermal deterioration is essentially an oxidation chemical reaction, chemical bonds break due to the thermally induced vibration of the chemical bonds. Macroscopically, the insulation is more brittle, has lower mechanical strength and less capability to bond the tape layers together. In thermoplastic insulation temperature above critical threshold causes flowing and make insulation softer [3].

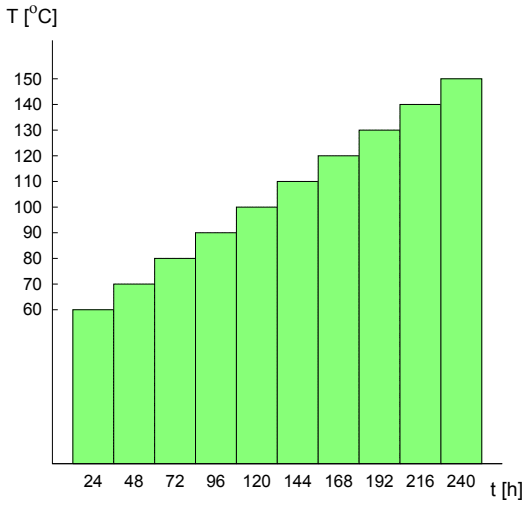


Fig. 1. Graph of short-term thermal aging.

In order to evaluate short-term thermal degradation on insulation system were made five measurements before thermal stress and five measurements after thermal stress. Phase resolved partial discharge pattern (histograms) were created. Compare given parameters of depicted phase resolved partial discharge patterns is possible to acquire information about condition of the insulation system.

II. Experimental setup

The short-term thermal aging was done on high voltage coil with the rated voltage U_n = 6 kV and with the rated power P_n = $0,2$ MW. Insulation of measured coil was $Samicatherm$ 366.28 - created by micacalcined tape amalgamated by epoxy asphalt by means of Resin Rich technology without semiconducting coating protection on the end of the slot part of coil. The coil was placed in drying-oven MLW TS 400 where temperature changed from 60 °C to 150 °C with step 10 °C and with time duration 24 hours per one temperature (see Fig.1). It meanse the coil was overall thermal stressed 240 hours with different temperature.

Partial discharges measurement was done according to IEC 60270 (off-line method with direct galvanic coupling see Fig.2) in laboratory conditions. Coil hung on insulation rope placed in drying-oven. The end of coil was galvanic joined

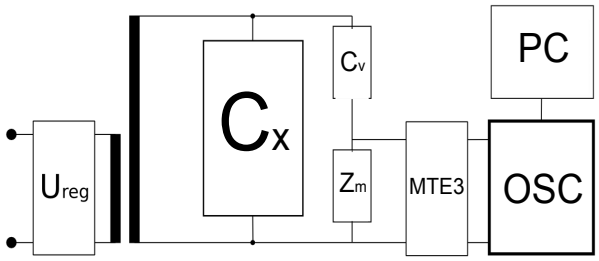


Fig. 2. Scheme of experimental setup

by homogeneous electrode to eliminate the corona discharges rose from harsh protrusions. In order to model of putting coil in the stator slot, the slot part of coil was grounded by means of galvanic network wrapped by conductive foil in the middle of coil's slot part.

For reaching reliable results coil was measured five times before and after short-term thermal stress.

- OSC - oscillosope
- C_v - coupling capacitor
- U_{reg} - voltage regulator
- PC - personal computer
- Z_m - measured impedance
- C_x - measured object (coils)
- $MTE3$ - measured equipment

A. *Teoretical analysis*

Data from partial discharge measurements were acquired by Agilent program which process its and safe in five different files. Two files contain information about pulse magnitude of apparent charge and pulse count in given phase window. In the next process data were calculated in Octave programme. The phase resolved partial discharges pattern - the maximum pulse height distribution $H_{qmax}(\varphi)$, the mean pulse height distribution $H_{qn}(\varphi)$ and the mean pulse count distribution $H_n(\varphi)$ were generated and depicted in two dimensional plots (see Fig.3 - Fig.5).

These histograms depicted partial dischrge activity in phase windows. The maximum pulse height distribution $H_{qmax}(\varphi)$ is maximal dicharge activity of captured apparent charges in individual phase windows by means of mathematical procedure:

$$H_{qmax,i} = \max_{0 \le j \le n} [q_{i,j}] \tag{1}$$

Index i denotes the phase window, index j denotes the corresponding value of apparent charge in the i-phase window and n is the number of captured periods and it was varied from $n = 440 - 920$ periods.

The mean pulse height distribution $H_{qn}(\varphi)$ is discharge activity in individual phase windows where was calculated the mean value of captured magnitues of apparent charges. The mean value means fraction of sum apparent charges in particular phase windows to frequencies in the same phase windows. Mathematical description:

$$H_{qn,i} = \frac{1}{n_i} \sum_{j=1}^{n} q_{i,j} n_{i,j} \tag{2}$$

where n_i denotes the total frequency of discharges in the phase window and $\sum n_{i,j} = n_i$, n is the number of captured periods

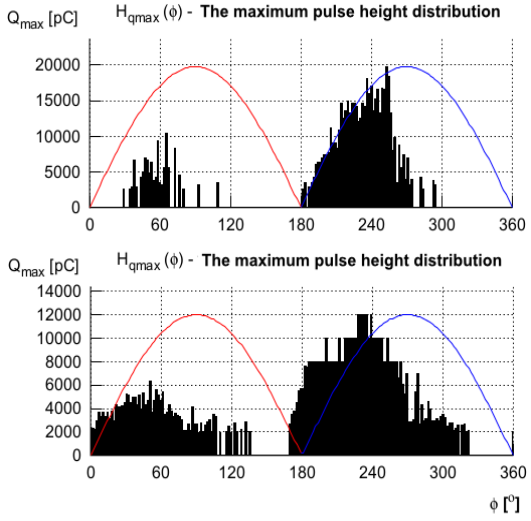


Fig. 3. Graphs of distributions of $H_{qmax}(\varphi)$ at 6kV before and after short-term thermal stress.

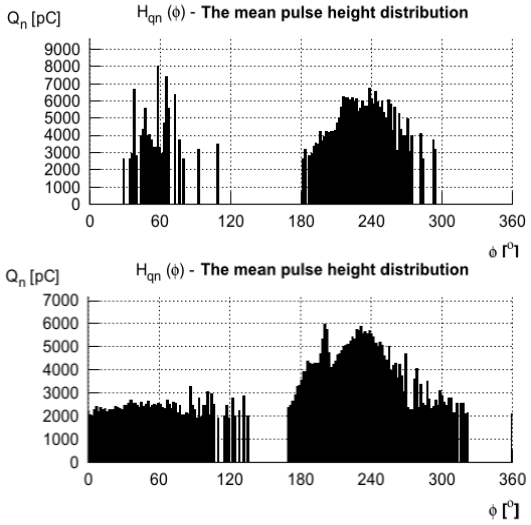


Fig. 4. Graphs of distributions of $H_{qn}(\varphi)$ at 6kV before and after short-term thermal stress.

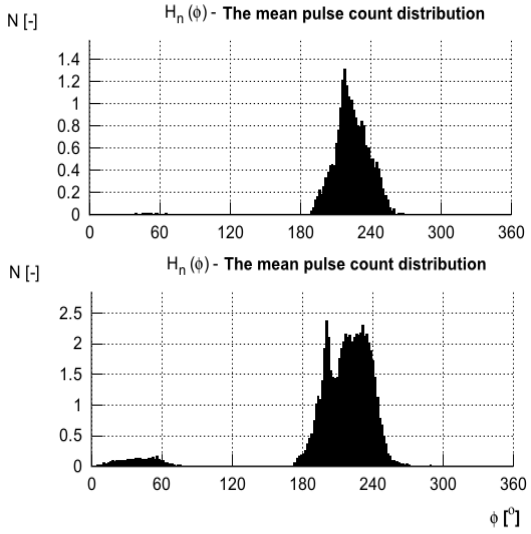


Fig. 5. Graphs of distributions of $H_n(\varphi)$ at 6kV before and after short-term thermal stress.

of the applied voltage same as in previous case $n = 440 - 920$.

TABLE I
TABLE OF MEASURED VALUES FOR INCEPTION VOLTAGE U_i BEFORE AND AFTER SHORT-TERM THERMAL STRESS

Before thermal stress				
U $[kV]$	Q_{max} $[pC]$	N $[-]$	φ^+ [°]	φ^- [°]
2,6	400	0,12	50 - 70	230 - 250
3,2	1200	1,7	40 - 70	210 - 250
2,6	350	0,10	40 - 70	220 - 260
3,2	1600	0,04	30 - 70	230 - 250
2,8	350	0,35	50 - 70	240 - 260
After thermal stress				
U $[kV]$	Q_{max} $[pC]$	N $[-]$	φ^+ [°]	φ^- [°]
3,0	450	0,14	30 - 70	210 - 260
3,2	1200	0,11	30 - 70	210 - 250
3,2	1400	0,16	20 - 100	220 - 260
3,2	800	0,03	-	230 - 250
3,0	400	0,03	10 - 40	240 - 260

Indexes i, j have the same meaning as in Equation 1.

The mean pulse count distribution $H_n(\varphi)$ represents a vector of the mean values of frequencies in individual phase window:

$$H_{n,i} = \frac{1}{n_i} \sum_{j=1}^{n} n_{i,j} = \frac{n_i}{n} \tag{3}$$

n_i and n have the same meaning as in the equation 2.

III. REVIEWED OF RESULTS

The main goal of this article was monitoring diferences between partial discharge activity before and after short-term thermal stress and reveal how temperature influence partial dischrges activities. Comparison of partial discharge activities were done by means of three voltage values:

- *The inception voltage value* U_i - the voltage value in which was captured stabil discharge activity,
- *The rated phase-to-ground voltage value* Un *of coil* - the voltage value 3,6kV,
- *The rated phase-to-phase voltage value* Uz *of coil* - the voltage value 6kV.

Beside these voltage values were ascertained magnitudes of the maximum pulse height distribution $H_{qmax}(\varphi)$, magnitude of the mean pulse count distribution $H_n(\varphi)$ and interval of angles in which partial discharge originated. There are values from individuals measurements in the next tables (see Tab. I - Tab. III).

The Tab. I contian measured values in term of inception voltage U_i. As can be seen the lower inception voltage values were in measurements before thermal stress. The values were from range of 2,6 - 3,2 kV. After thermal stress stabil discharge activity was captured above voltage 3,0 kV inclusive. The maximal values of apparent charges before stress were in range of 400 - 1600 pC with the mean value 780 pC. After thermal stress values were in range of 450 - 1400 with the mean value 850 pC but with lower pulse count.

The Tab. II show measured values at the rated phase-to-ground voltage U_n of coil 3,6 kV. Before stress were measured maximal values of apparent charges in range of 1200 - 4000 pC with the mean value 4050 pC and in case after thermal stress in range of 3500 - 12000 pC with the mean value 6100pC. As you can see the apparent charge were higher in the case of measurement after thermal stress.

TABLE II
TABLE OF MEASURED VALUES FOR PHASE-TO-GROUND VOLTAGE U_n BEFORE AND AFTER SHORT-TERM THERMAL STRESS

Before thermal stress				
U $[kV]$	Q_{max} $[pC]$	N $[-]$	φ^+ [°]	φ^- [°]
3,6	1200	0,01	50 - 70	230 - 270
3,6	4000	0,50	40 - 70	200 - 270
3,6	4000	0,05	-	210 - 300
3,6	3500	0,05	30 - 70	220 - 290
3,6	3500	0,16	-	240 - 260
After thermal stress				
U $[kV]$	Q_{max} $[pC]$	N $[-]$	φ^+ [°]	φ^- [°]
3,6	12000	0,40	30 - 40	200 - 300
3,6	4500	0,35	30 - 70	230 - 270
3,6	4500	0,40	30 - 80	220 - 300
3,6	6000	0,04	-	220 - 250
3,6	3500	0,27	30 - 70	230 - 270

TABLE III
TABLE OF MEASURED VALUES FOR PHASE-TO-PHASE VOLTAGE U_z BEFORE AND AFTER SHORT-TERM THERMAL STRESS

Before thermal stress				
6,0	14000	0,12	50 - 70	210 - 260
6,0	18000	0,80	40 - 70	180 - 260
6,0	25000	0,90	20 - 80	180 - 290
6,0	18000	0,50	30 - 60	180 - 260
6,0	19000	0,35	30 - 110	180 - 290
After thermal stress				
U $[kV]$	Q_{max} $[pC]$	N $[-]$	φ^+ [°]	φ^- [°]
6,0	23000	1,00	10 - 120	190 - 290
6,0	8000	1,50	10 - 130	190 - 300
6,0	25000	1,40	20 - 70	180 - 290
6,0	10000	0,03	-	230 - 250
3,0	12000	2,30	0 - 130	180 - 320

The Tab. III show measured values at the rated phase-to-phase voltage U_z of coil 6,0 kV. In first case before thermal stress maximal values of apparent charge were in range of 14000 - 25000 pC with the mean value 188000 pC. After thermal stress magnitudes of apparent chrages were in range of 8000 - 25000 pC with the mean value 15600 pC.

It is a surface discharge activity at both voltage values U_n and U_z. In first case before stress the magnitudes of apparent charges were lower at U_n than in the case after thermal stress. But at U_z maximal values of apparent charges droped down after thermal stress.

IV. CONCLUSION

The aim of this article was monitoring short-term thermal degradation on the high voltage stator coil without corona protection. The results showed that thermal stress increase

inception voltage U_i (stabil partial dischrage activity). The lower inception voltage U_i hint on more enumerable inner dischrage activity. The insulation after thermal stress become more baked and more compact, so the number of inner cavities decrease. The another hands showed that thermal stress increase partial discharge activity at the rated phase-to-ground voltage value U_n, but decrease partial discharge activity at the phase-to-phase voltage value U_z. Whole influence of short-term thermal stress with time duration 240 hours is positive, insulation become more compact

ACKNOWLEDGMENT

This work was supported by Agency of the Ministry of Education of the Slovak Republic for the Structural Funds of the EU under the project Development of Low Power Static Supply for Electric Systems (project number: 26220220029, priority axis 2 Support to research and development).

REFERENCES

[1] B. Yazici, "Statistical pattern analysis of partial discharge measurements for quality assessment of insulation systems in high voltage electrical machinery," *Symposium on Diagnostics for Electric Machines, Power Electronics and Driver*, pp. 24–26, 2003. [Online]. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=01360005

[2] S. M. B. R. Sahoo, N.C., "Trends in partial discharge pattern classification: A survey. sahoo et.al: Trends in partial discharge pattern classification," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 2, pp. 248–264, 2005.

[3] G. Stone, "Partial discharge diagnostics and electrical equipment insulation condition assessment," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 12, pp. 891–903, 2005. [Online]. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=123443

Influence of Temperature on Dielectric Spectroscopy of Magnetic Fluid Based on Transformer Oil

[1]*Jozef Király (2nd year),* [2]*Martin German-Sobek (2nd year)*
Supervisor: [3]*Roman Cimbala*

[123] Dept. of Electrical Power Engineering, FEI TU of Košice, Slovak Republic

[1]jozef.kiraly@tuke.sk, [2]martin.german-sobek@tuke.sk, [3]roman.cimbala@tuke.sk

***Abstract*—The article deals with magnetic fluid based on a transformer oil and dependence of capacitance and dissipation factor tgδ by two different temperatures. It also deals about influence of static external magnetic field acting to the behavior of these parameters.**

***Keywords*—Magnetic fluid, magnetite, thermal dependence, capacitance, dissipation factor**

I. Introduction

Magnetic fluids are colloid solutions based on carrier medium and ferromagnetic nanoparticles, which are coated with surface active substance which is also called surfactant. Ferromagnetic nanoparticles performs role of carrier of magnetic properties in the magnetic fluid. The function of surfactant in these fluids is to set up homogeneously dispersion of nanoparticles in volume of fluid. We can talk about stabile magnetic fluids in case, when concentration neither distribution of magnetic nanoparticles is unchanged. Thus composed magnetic fluids can form equivalent of many conventional fluids used in electric engineering, engineering but also in medicine with added magnetic properties. [1].

Particles dispersed in magnetic fluid reaches usually dimension about 3 - 15 nm. Typical content of these particles in fluid is 10^{23} m^3. Following to size of these particles their behavior is like a mono domain, thus every of them behave like an independent dipole with size $M_d=10^9 \mu B$ where μB is Bohr magneton, therefore the smallest amount of electron magnetic moment. [1].

When we ignore interaction between nanoparticles, we can predict, that magnetic fluid will behave like a paramagnetic fluid, so these behavior we can describe by Langevin function. [2]

In case when magnetic moments are bigger than magnetic moments of atoms of usually paramagnetic substances, we can talk about a superparamagnetic behavior. [1].

II. Magnetic Fluids for Power Transformers

For production magnetic fluid for usage in high voltage devices it's possible to use mineral oil enriched with inhibitors. The role of surfactant performs oleic acid. These oleic acids are organic acids, which are in group of higher olefin-carbon acids. [1]. Molecules of oleic acid are amphiphile. It means that they have a polar hydrophilic head and nonpolar hydrophobic part. This hydrophobic part is acid formed with long chain in case of this. Polar head will snap this long chain to magnetite nanoparticle, resulting in solvate cover. The production of magnetic fluid is possible by process of precipitation, i.e. shrinking of soluble substance from solid dilution. This process was used also for preparing of tested fluids. [1]

III. Magneto-Dielectric Anisotropy in Magnetic Fluids

Magnetodielectric anisotropy can be described as influence of external magnetic field on to dielectric properties of magnetic fluids. These properties are affected by changes in magnetic fluid structure, due to processes of magnetic nanoparticles structuralisation. Structuralisation of magnetite nanoparticles into direction of acting field (magnetic field, electric field) can be described by processes of magnetophoresis and dielectrophoresis. Also influence of temperature is not inconsiderable due to phenomenon of thermomagnetic convection and as well a thermal dependence of magnetization according to Langevin function. These structural changes cause that dielectric parameters (capacitance, dissipation factor, electric strength, dielectric constant etc.) are changing. Process of creating long chains due to magnetophoresis, dielectrophoresis and thermomagnetic convection is called bridging.

A. Magnetophoresis

One type of mechanism of creating bridges is bridging due to presence of magnetic field. General principles of magnetophoresis can be briefly explained by few equations. Potential energy of a particle can be written as [3]:

$$U = -\frac{(\chi_p - \chi_m)}{2\mu_0} V B^2 \quad (1)$$

Where:
V - Volume of particle
χ_p - Magnetic susceptibility of particle
χ_m - Magnetic susceptibility of medium
μ_0 – Magnetic permeability of vacuum
B – Magnetic flux density

Equation for magnetic force, which is acting on particle, can be written as [3]:

$$\boldsymbol{F} = -gradU = \frac{(\chi_p - \chi_m)}{2\mu_0} V(\boldsymbol{B}\cdot\nabla)\cdot\boldsymbol{B} \quad (2)$$

Due to (2) an inhomogeneous magnetic field can create a force, which can act on particle in liquid. The drag force which acts on particle during a movement can be described as [3]:

$$\boldsymbol{F_D} = -6\pi\eta r\boldsymbol{v} \quad (3)$$

Influence of environment viscosity is given in (3) by η. Thus force acting is directly proportional to viscosity of environment and to radius of particle. [3]

Equation for magnetophoretic velocity can be written as:

$$\boldsymbol{v} = \frac{2}{9}\frac{(\chi_p - \chi_m)}{\mu_0\eta} r^2(\boldsymbol{B}\cdot\nabla)\cdot\boldsymbol{B} \quad (4)$$

This equation shows that velocity of particle is directly proportional to difference between magnetic susceptibility of medium and particle. [3]

Magnetophoresis can be in general divided into two types. Positive magnetophoresis appears in case, when permeability of environment (fluid) is lower than permeability of particles. In this case are particles attracted to local maxima of inhomogeneous magnetic field and repelled from minima. For negative magnetophoresis is necessary, that permeability of environment is higher than permeability of particles. [4]

B. *Dielectrophoresis*

Whereas magnetophoresis occurs under influence of magnetic field, for dielectrophoresis is necessary presence of electric field. This phenomenon can be short described as force acting on particle during exposition of inhomogeneous electric field.

For presence of dielectrophoresis is required presence of inhomogeneous electric field. This field creates a gradient which causes migration of dipole particles. It should be noted, that particles must be dipoles, but they need not have charge. Degree of dielectrophoresis and thus velocity of particles migration is dependent on particle size and shape, properties of environment and on properties of acting field. [5]

Equation for force acting on dipoles without charge can be written as [5]:

$$F = \frac{\pi\varepsilon_m\varepsilon_0}{12} D^2 L\left[\frac{\alpha}{\alpha-1} - f(\beta)\right]^{-1} \nabla E^2 \quad (5)$$

In this equation (5) is relative permittivity of environment represented by ε_m. Equation for this model assumes that particles are prolate spheroids with diameter D and length L. Ratio between relative permittivity of particles and environment is represented by α. Aspect ratio between diameter and length is represented by β.[2]

Migration of particles which is parallel to drag force is given by:

$$F = 3\pi\eta D v g(\beta) \quad (6)$$

For conductive particles we can assume that $\varepsilon_f \to \infty$, therefore $\alpha\to\infty$ and $\frac{\alpha}{\alpha-1} \to 1$. After combining equations (5), (6) and substituting $\frac{\alpha}{\alpha-1} \to 1$, and taking the limit $\beta\to\infty$ we can write equation for particle velocity [5]:

$$v = \lim_{\beta\to\infty} \frac{\varepsilon_m\varepsilon_0}{24\eta} L^2 \frac{\ln 2\beta - 0{,}5}{\ln 2\beta - 1} \nabla E^2 \quad (7)$$

The influence of particle aspect ratio is provided only for β>5. Equally, β must be for this approximation >2. Velocity is directly proportional to square of the length. [5]

C. *Bridging of nanoparticles*

For bridging, hence for creation of long chains is necessary sufficient long time of magnetic field presence. Experimentally was found, that for origin of long chain with length of 100 μm and more is necessary time approx. 100 s. This fact was considered for all experiments

In case of usage of these fluids as coolant in power transformer, it must be mentioned, that in this case, except presence of magnetophoresis and dielectrophoresis, is there also present phenomenon of thermomagnetic convection. This phenomenon can cause migration of nanoparticles and thus increases possibility of creating bridges in volume of magnetic fluid. Condition for this phenomenon is presence of temperature gradient and magnetic field. For use in cooling system of power transformer are conditions for thermomagnetic convection satisfied. The source of temperature gradient is heat produced by AC current in windings. Thus heat generated by winding and core of transformer creates in area inhomogeneous magnetic field and temperature gradient.

It has to be mentioned, that magnetization of magnetic fluid is dependent on the value of magnetic field intensity and as well as on the carrier medium susceptibility. In a magnetic fluid where are present varying temperatures, the susceptibility is a function of the temperature as is showed on figure 1. Due to these facts, is aim of this article measurement of selected dielectric parameters by two different temperatures. [2]

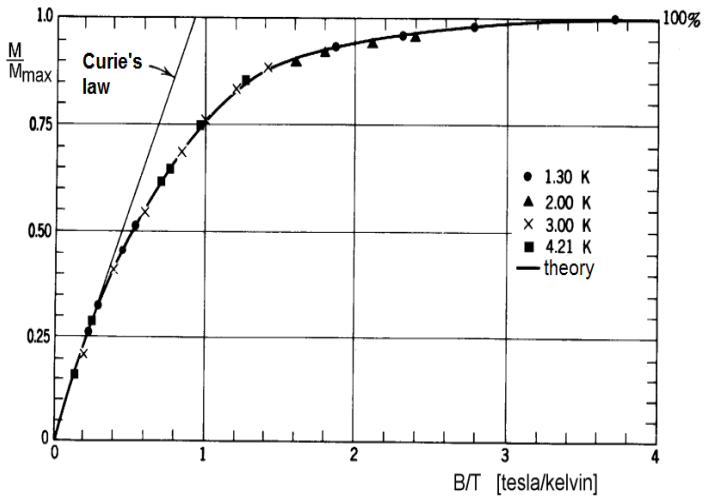


Fig. 1. Langevin function for dependence of M/M_{max} and B/T

IV. EXPERIMENT

A. *Experimental Base*

Experiment was performed on sample of magnetic fluid placed in vessel made from PTFE in which was placed electrode system composite from Rogowski electrode system. External applied magnetic field was created by pair of permanent NdFeB magnets with magnetic induction 40 mT. Position of magnets was perpendicular or in axis of electrode system. For comparison, measurement has been performed also without external magnets. Distance between electrodes was set for measurement to 1 mm. Distance between magnets was 5 cm and in every case was the electrode system placed in the middle.

The measurement was realized for two temperatures 20°C and 50°C. Magnetization of tested fluid (M=1mT) was given by manufacturer. Assignment of capacitance and dissipation factor was made by LCR meter in frequency range from 10 Hz to 2 MHz.

B. *Measurement of Capacitance and Dissipation Factor*

As is shown on figures 2 and 3, anisotropy of capacitance is strongest in case of acting of perpendicular magnetic field. This is caused by fact that magnetophoretic force is stronger than dielectrophoretic force, which is acting in direction of electrode axis. Thus this force is creating a long chain which affects dielectric constant, which is proportional to capacitance. In case of dissipation factor is anisotropy also observable, mainly in frequency range up to 1000 Hz.

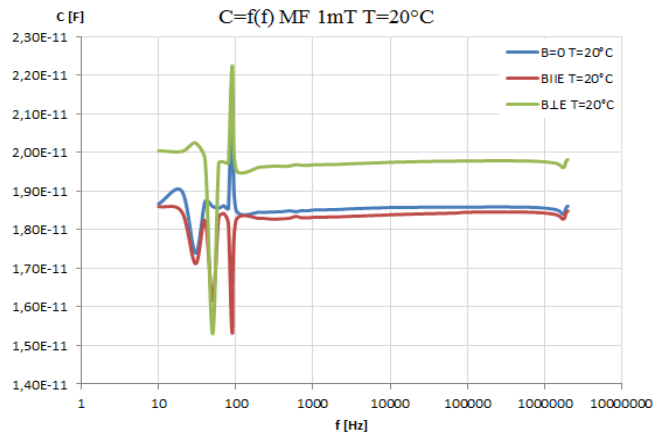


Fig. 2. C=f(f) for MF with magnetization 1 mT by temperature 20°C

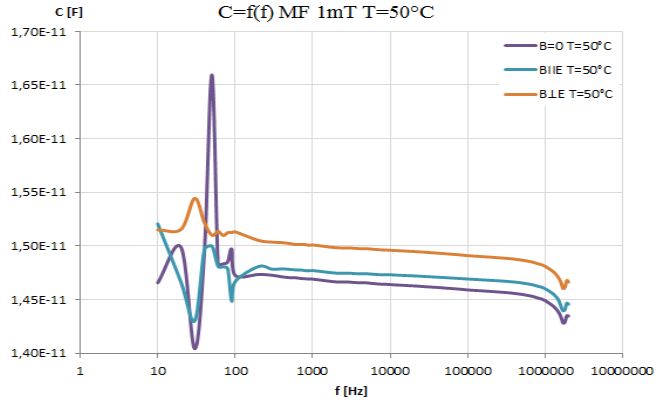


Fig. 3. C=f(f) for MF with magnetization 1 mT by temperature 50°C

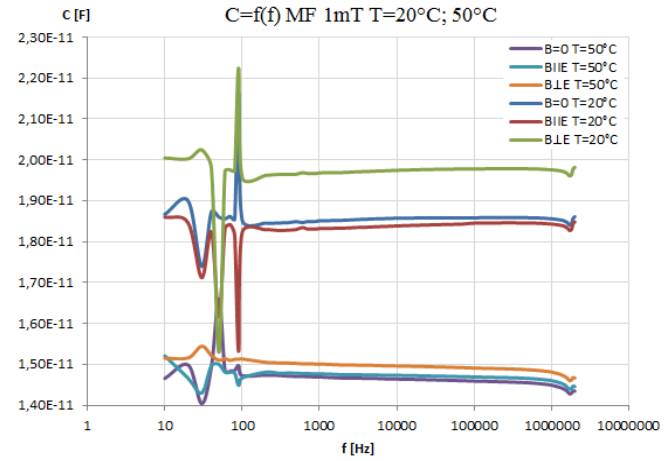


Fig. 4. Comparison of C=f(f) dependencies for MF with magnetization 1 mT by temperature 20°C and 50°C

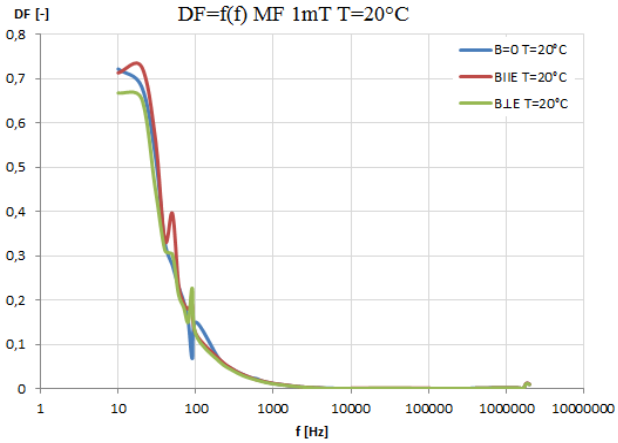


Fig. 5. DF=f(f) for MF with magnetization 1 mT by temperature 20°C

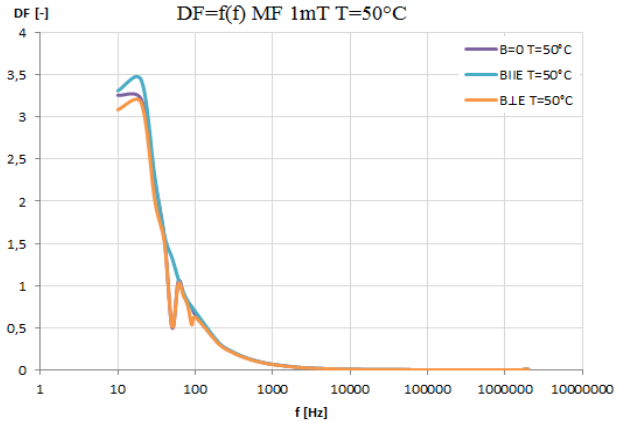


Fig. 6. DF=f(f) for MF with magnetization 1 mT by temperature 50°C

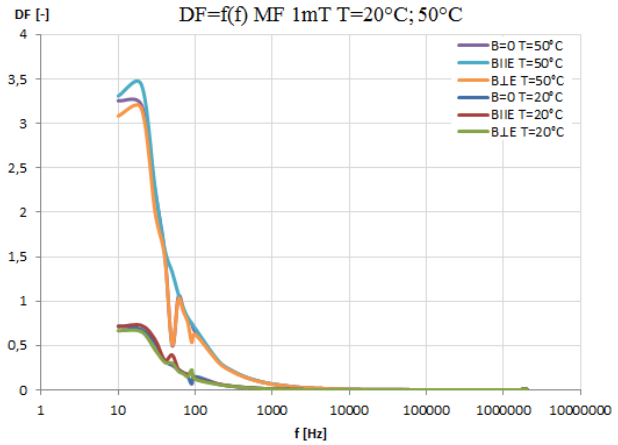


Fig. 7. Comparison of DF=f(f) dependencies for MF with magnetization 1 mT by temperature 20°C and 50°C

V. Conclusion

This article was aimed to perform comparison of magnetodielectric anisotropy by different temperature. This anisotropy was shown on measurement of capacitance and dissipation factor. The effect of influence of external magnetic field is best recognizable by measurement of capacitance. These measurements were realized for demonstration of differences between pure transformer oil and magnetic fluid based on transformer oil.

Significant differences between values without external magnetic field and with perpendicular magnetic field are caused by bridging phenomenon between electrodes. As was mentioned in article, this phenomenon is caused by magnetophoresis, dielectrophoresis and also thermomagnetic convection.

Influence of temperature to magnetodielectric anisotropy is less significant, because of similar values measured by lower and higher temperatures. Higher temperature affects mainly values of capacitance and dissipation factor. Character of anisotropy was not significantly affected.

Acknowledgment

We support research activities in Slovakia / Project is cofinanced from EU funds. This paper was developed within the Project "Centrum excelentnosti integrovaného výskumu a využitia progresívnych materiálov a technológií v oblasti automobilovej elektroniky", ITMS 26220120055.

This work was supported by scientific Grant Agency of the ministry of Education of the Slovak Republic project VEGA No. 1/0487/12.

References

[1] HERCHL, František: Štúdium elektrofyzikálnych vlastností magnetických kvapalín na báze transformátorového oleja: Dizertačná práca. Košice SAV ÚEF, 2008. 87 p.

[2] TIRPÁK, Andrej: Elektromagnetizmus. Bratislava: Univerzita Komenského, Fakulte matematiky, fyziky a informatiky, Katedra rádiofyziky, 2004. 711 p. ISBN 80-88780-26-8

[3] WATARAI, Hitoshi, SUWA a Yoshinori IIGUNI. Magnetophoresis and electromagnetophoresis of microparticles in liquids. Analytical and Bioanalytical Chemistry. 2003, s. 1693-1699. ISSN 1618-2642. DOI: 10.1007/s00216-003-2354-7.

[4] GAO, Y, JIAN, L.F. ZHANG a J.P. HUANG. Magnetophoresis of Nonmagnetic Particles in Ferrofluids. In: The Journal of Physical Chemistry C. 2007, s. 10785-10791. ISSN 1932-7447

[5] Mahmud, S.; Golosnoy, I.O.; Chen, G.; Wilson, G.; Jarman, P.: "Numerical simulations of bridging phenomena in contaminated transformer oil," Electrical Insulation and Dielectric Phenomena (CEIDP), 2012 Annual Report Conference, pp.383-386, 14-17 Oct. 2012 doi: 10.1109/CEIDP.2012.6378800

Iterative detection of coded SC-FDMA symbols with hard and soft decision detector

Denis Dupák (3st year), Juraj Gazda
Supervisor: Dušan Kocur

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

denis.dupak@tuke.sk

***Abstract*—This paper investigates the application of soft decision detector in cooperation with iterative reciever in the nonlinearly distorted Long Term Evolution (LTE) uplink. In general, the soft decision decoding of convolutional codes provide performance improvement of transmission system and it is already adopted in GSM systems. However, its application in the LTE standard has not been studied yet. Therefore, the motivation of this paper is to execute on the soft decision performance in the LTE uplink and to provide the comparison with conventional hard decision detector. We will focus on 16(64)-QAM baseband modulation scheme in Single Carrier Frequency Division Multiplexing (SC-FDMA) system. The performance analysis show the considerable performance improvements of presented technique performing over frequency selective channels and additive white Gaussian noise (AWGN) channels.**

***Keywords*—SC-FDMA, OFDM, BER, Peak-to-Average Power Ratio, Soft and Hard decision.**

I. Introduction

The demand for high data rates transmission over multipath radio channels has increased rapidly. To fulfill this requirement, utilization of multi-carrier based transmission techniques seems to be an inevitable solution. The major candidate for 3G and 4G wireless communication systems with very promising potencial is Orthogonal Frequency Division Multiplexing (OFDM) scheme. The OFDM is spectrally efficient modulation technique that has recently gained much popularity due to high speed and very effective data trasmission over multipath fading channels. OFDM based transmission systems are characterized by large number of benefits in comparison with traditional schemes. This makes it superior compared to previously introduced wireless standards and therefore, it is expected its wide deployment in forthcoming years. Moreover, further research and engineering activities leading to the standard performance improvement are still underway. This new standard is recognized as LTE ADVANCED and it is anticipated to be fully operational in future.

LTE and LTE ADVANCED utilize in the downlink OFDM scheme, which is in general extremely robust against multipath propagation and allows using very simple one-tap channel equalization. On the other hand, high Peak-to-average power ratio (PAPR) of OFDM signal makes it very sensitive to the nonlinear amplification which results in high Bit Error Rate (BER) penalty as well as to the enormous out-ofband radiation. These effects have harmful impact on the overall OFDM transmission system performance and therefore strict requirements mitigating these effects must be taken.

Large sensitivity to nonlinear amplification has greatly limited the practical applications of OFDM transmission systems. Therefore, in order to reduce the effect of nonlinear amplification, the new modulation scheme, Single Carrier Frequency Division Multiplexing (SC-FDMA), was presented in LTE uplink [1]. SC-FDMA is characterized by significant lower PAPR compared to that of OFDM, nevertheless, it is still higher compared to the conventional single carrier systems.

In order to alleviate the effect of nonlinear amplification in SC-FDMA transmission systems, many approaches based on different techniques have been introduced. Frequently used solution in the transmitter is to back-off the operating point of nonlinear amplifier, but this approach results in significant power efficiency penalty. Alternative approaches for SC-FDMA performance improvement are realized by applications of other usually computational demanded PAPR reduction methods at the transmitter side, e.g. active constellation extension [2], tone reservation [3] or selected mapping [4].

Another well known and promising solution to reduce BER of nonlinearly distorted multicarrier systems is to use nonlinear detection at the receiver side. The first contribution on this topic was proposed by Kim and Stuber in [5]. This technique reduces the clipping noise of OFDM symbols by decisionaided reconstruction at the receiver. In [6] Declerq et al. proposed reducing the clipping noise in OFDM by introducing a Bayesian interference to the received signal. Finally Chen et al. [7] and Tellado et al. [8], proposed iterative techniques to estimate and eliminate the clipping noise in OFDM.

As it was shown in [9], the performance of SC-FDMA transmission systems was improved by using of iterative nonlinear detection at the receiver side. We made a modification of presented scheme for coded SC-FDMA transmission. This modification is based on jointly using the original receiver structure presented in [9] concatenated with the channel decoder represented by the hard and soft output Viterbi Algorithm.

In this paper, we will focus to investigate the performance of SC-FDMA transmission systems with iterative nonlinear receiver and hard decision detector, which is commonly used in [7], [8], [9] and the same SC-FDMA transmission system with soft decision detector. It will be showed that soft decision

SC-FDMA transmission system performs well and overcome the performance of the conventional hard decision SC-FDMA system in both investigated criteria BER, especially if strong nonlinear distortion due to High Power Amplifier (HPA) is present. Also, we have to mention that computation complexity increases with using of soft decision detector, but it is small price for performance gain about 3dB.

II. Model of SC-FDMA Transmission System

The SC-FDMA transmitter transforms the sequence of binary bits to the sequence of modulated subcarriers. To achieve this, it performs similar signal processing operation than that of OFDM. Simplified block scheme of SC-FDMA transmission is given in Fig. 1.

In SC-FDMA, the bit sequence is mapped onto *N* complex modulation symbols in the first step. Commonly used baseband modulation schemes in upcoming LTE standard

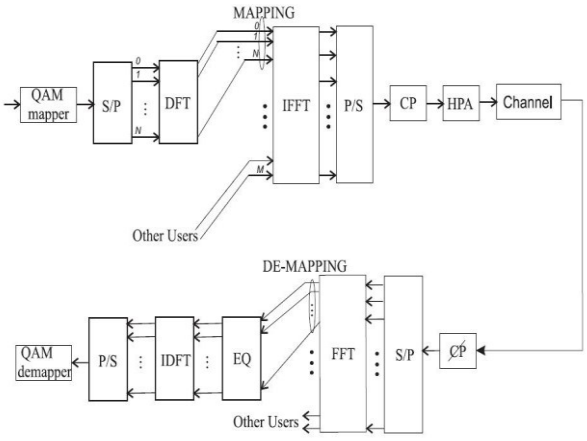


Fig. 1. Block scheme of SC-FDMA transmission system

include QPSK, 16-QAM and 64-QAM. In general, transmitter adopts the modulation scheme to match the particular channel conditions and characteristics for the certain time instance. Afterwards, block of *N* data symbols are applied to a size *N* discrete Fourier Transform (DFT). This operation is in the literature also described as a DFT-precoding operation. Application of DFT-precoding at the transmitter side is a promising solution for the reduction of the envelope fluctuation of the original OFDM signal. The next stage of SC-FDMA procedure is to shift the baseband DFT representation of the time-domain SC-FDMA symbol to the desired part of the overall channel bandwidth. Since the signal is now represented as a DFT, assigning individual subcarriers to the user is only matter of frequency shifting and copying the *N* bins into a larger DFT *M* bins space that can be up to the size of the system channel bandwidth. In LFDMA, each user terminal adopts consecutive adjacent subcarriers for the transmission. As a results of this fact, particular mobile terminal use only a fraction of the overall bandwidth and only a small fraction of frequency diversity is used. This fact is a special relevance for the systems, where channel state information is perfectly know for the each time instant, hence no further kinds of diversity is necessary to use.

Let a_i, $i = 0, \ldots, N-1$ be the complex data symbols, then the signal at the output of DFT-precoder can be expressed as

$$S_k = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} a_i e^{-j2\pi ik/N},\ k = 0,...,N-1. \quad (1)$$

Consider a baseband OFDM symbol *s(t)* defined over the time interval $t \in [0, T_S)$,

$$s(t) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} S_k e^{j2\pi(k+k_0)t/T_S},\ k = 0,...,N-1, \quad (2)$$

where k_0 is the position of the first assigned subcarrier. For the sake of brevity and without loss of generality we assume $k_0 = 0$. If *s(t)* is sampled at a frequency LN/T_s, where $L = M/N$ is the oversampling factor and N/T_s is the Nyquist rate, the signal at the output of the SC-FDMA modulator is

$$s_n = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} S_k e^{j2\pi kn/M},\ k = 0,...,M-1. \quad (3)$$

By substituting (1) in (3), the SC-FDMA signal is found to be

$$s_n = \frac{1}{N} \sum_{i=0}^{N-1} a_i \sum_{k=0}^{N-1} e^{j2\pi k(n-Li)/M}. \quad (4)$$

By analyzing (4) at multiples of the oversampling factor, $n = Lr$, one observes that the L_r-th sample of the time domain SC-FDMA signal is equal to the data symbol a_r. The samples at positions $n \neq Lr$ describe the transition of the time domain signal between values a_r and a_{r+1}. The presence of these transitions between modulated symbols increase PAPR and cubic metric (CM).

III. Soft Decision Detector

Let m_i denote the mean valuc of *x* (demodulator output) for the *i*-th bit. We will assume, without loss of generality, that things are normalized to make m_i +1 for a binary 0 and -1 for a binary 1 when no fading occurs on the channel. Let v_i be + 1 or -1 when the ith bit transmitted is a binary 0 or 1, respectively. Let w_i denote a real coefficient which is indicative of the signal gain variations due to fading at the instant the *i*-th bit is transmitted on the channel. Then, the output of the demodulator corresponding to the ith bit may be expressed as

$$x_i = v_i w_i + n_i, \quad (5)$$

in which n_i denotes the additive white Gaussian noise at the output of the receiver.

With the noise being Gaussian, the probability density of x_i given that the ith bit transmitted is v_i takes the familiar form

$$p(x_i | v_i) = \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left((x_i - v_i w_i) / \sigma_i \sqrt{2} \right)^2, \quad (6)$$

where the σ_i^2 presents the noise variance (i.e., σ_i^2 s proportional to the average power of the noise) at the instant the *i*th bit istransmitted on the channel. It is noted that this parametermay vary over the bit sequence if the noise is not

stationary. Since the noise is assumed to be independent from bit to bit,l which stems from the white nature assumed for it, the likelihood function L for a path consisting of a sequence of *N* hits within the trellis in the Viterbi decoding process will be simply the product of the probability functions expressed by (14) for each of the bits; that is,

$$L = \prod_{i=1}^{i=N} p(x_i \,|\, v_i) = \prod_{i=1}^{i=N} \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left(\frac{-\left(x_i - v_i w_i\right)}{\sigma_i \sqrt{2}} \right)^2 , \quad (7)$$

The optimally decoded sequence is then the sequence associated with the path in the trellis which maximizes the above likelihood function. Since the logarithmic is a monotonically increasing function of its argument, the optimum sequence would also be the sequence which would maximize the log of the likelihood function. Therefore, we can replace the likelihood function given in (14) with its natural log in the maximization process. This simplifies things in the sense of turning the product into a summation as follows:

$$l\log(L) = \sum_{i=1}^{i=N} \left[l\log\left(\frac{1}{\sigma_i \sqrt{2\pi}} \right) \exp\left(\frac{-\left(x_i - v_i w_i\right)}{\sigma_i \sqrt{2}} \right) \right]^2 . \quad (8)$$

Based on the above given likelihood probabilities the soft decision metrics is evaluated and the most likely path in the trellis is choosen in the soft decision detector.

IV. Simulation Results

In this section, performance improvement capabilities of soft decison technique are shown. The soft decision decoding technique was used to decode convolutional codes in convencial transmission systems and provided better performance compared to hard decision technique. In this paper, it was investigated performance improvement of SC-FDMA transmission system using soft decision detector in cooperation with iterative nonlinear estimation and cancellation of channel. The nonlinear iterative reciever used in our simulations was presented in [8]. As it was already shown in [10], nonlinear after certain point, the iterative detector does not provide the significant peformance improvement. Therefore there is no need for further iterations. To improve performace of iterative reciever, the hard detector used in reciever was replaced by the soft decision detector, proposed in previous section. The simulation results and comparision of transmission systems using soft and hard decisions are shown in figures below.

The performance results have been obtained using Monte Carlo computer simulations. The convolutional encoder with code rate *r=1/2* is used for channel coding. The simulations were performed for SC-FDMA system with 64 subcarriers with 16-QAM and 64-QAM modulation. The presented system operates at 2.4 GHz passband. As the propagation channel, typical 6-tap ITU Pedestrian A channel, which is corresponding to calling pedestrian in built-up area, and AWGN channel are used. To evaluate the nonlinear effects in SC-FDMA, nonlinear models of HPA are used. In presented simulaition results are used Soft limit and Saleh model of nonlinearity.

The operating point of the nonlinearity is defined by the so called input back-off (IBO) corresponding to the ratio between the saturated and average input powers.

The subject of our study focuses especially to soft decision BER reduction capabilities. The simulation results outlined in Fig. 2 corresponds to Soft limit model of nonlinearity with IBO level = 0dB and 16-QAM baseband modulator. SC-FDMA transmission system is operated over AWGN channel. As we can see on the figure, performance improvement of iterative method is negligibly for this scenario. Significant performance improvement (gain performance 1dB) is achieved using of soft decision technique.

Fig. 3 presents the simulation results for Saleh model of nonlinearity with IBO = 3dB, 64-QAM baseband modulator and AWGN channel is used. The transmission system using hard decision detector is marked by solid line and system using soft decision by dotted line. From the figure we can tell that using of soft decision detector provides significant BER performance improvement for all iterations and also for linear system.

Fig. 4 above presents the results for Saleh model of nonlinearity with IBO level = 3dB and 16-QAM baseband modulator. SC-FDMA transmission system is operated over ITU Pedestrian A channel. We investigated BER prerformance for E_b/N_0 from 22dB to 40dB. As we can see on the figure, performance improvement of soft decision detector is about 1dB for convencial and iterative reciever.

Fig. 5 presents the results for nonlinearly distorted SC-FDMA system with Saleh model of nonlinearity with IBO = 3 dB, 64-QAM baseband modulator. We used ITU Pedestrian A channel. In this scenario, we investigated BER prerformance for E_b/N_0 from 20dB to 45dB. From the figure we can see that using of soft decision detector provides the best BER performance improvement for non-iterative convential reciever (from 10^{-1} to 10^{-2} at E_b/N_0 = 45dB). Also significant performance improvement is achieved for iterative reicever. Gain performance for first, second iteration and linear system is 3db.

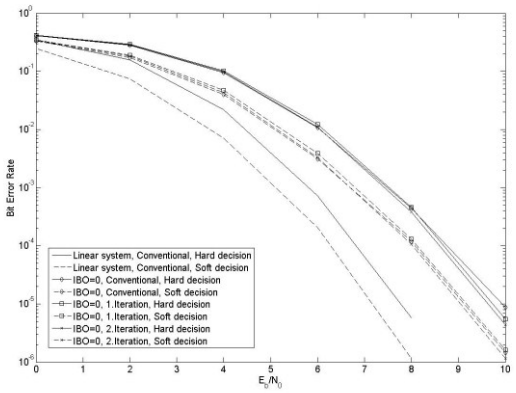


Fig. 2. Performance of the iterative algorithm with hard and soft decision detector for Soft limit nonlinearity model with IBO = 0, 64 subcarriers, 16 QAM, AWGN channel

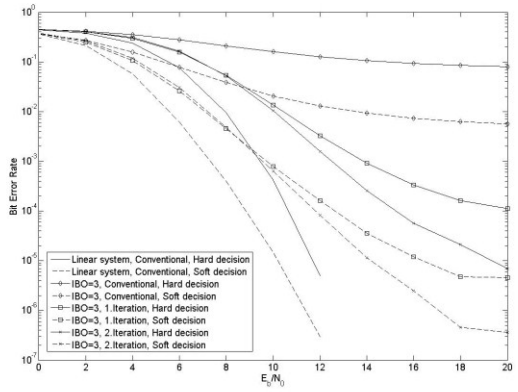


Fig. 3. Performance of the iterative algorithm with hard and soft decision detector for Saleh nonlinearity model with IBO = 3, 64 subcarriers, 64 QAM, AWGN channel l

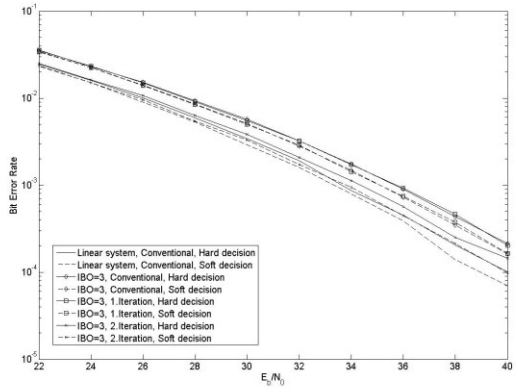


Fig. 4. Performance of the iterative algorithm with hard and soft decision detector for Saleh nonlinearity model with IBO = 3, 64 subcarriers, 16 QAM, ITU Pedestrian A channel

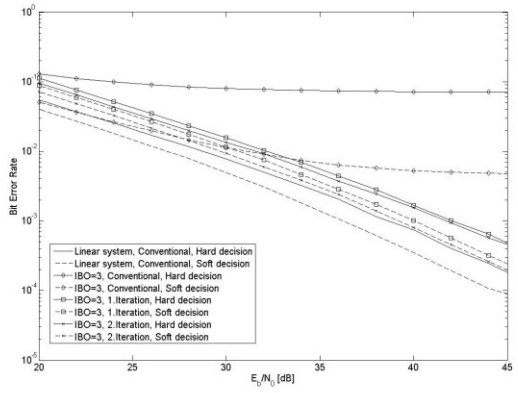


Fig. 5. Performance of the iterative algorithm with hard and soft decision detector for Saleh nonlinearity model with IBO = 3, 64 subcarriers, 64 QAM, ITU Pedestrian A channel

V. Conclusion

In this paper, we evaluated performance improvement of soft detection in the LTE uplink. The presented scheme combines the iterative receiver as proposed in [9],[10] with soft decision detector, in order to provide the better estimation of nonlinear distortion and thereby improving overall performance. It was shown that in particular scenarios, soft decision provide better BER performance compared to hard decision. However it should be noted that the performance gain is reached at the cost of the higher computation complexity.

Acknowledgment

This work is the result of the project implementation Short-Range UWB Sensor Networks for Detection, Localization and Tracking of Moving Persons (UWB-SeNet) (project VEGA 1/0563/13).

References

[1] H. G. Myung, D. J. Goodman, "A new air interface for Long Term Evolution", Winchester, UK, Wiley 2008.

[2] B.S. Krongold, D.L. Jones, "PAR Reduction in OFDM via Active Constellation Extension", *IEEE Transactions on Broadcasting*. vol. 49. no.3, pp. 258-268, Sep. 2003.

[3] M. Duemal, A. Behravan, T. Eriksson, J. L. Pijoan, "Evaluation of performance improvement capabilities of PAPR reducting methods", *Wireless Personal Communications*, vol. 47, pp.137-147, Oct. 2008.

[4] L. J. Cimini, N. R. Sollenberger, "Peak-to-Average Power Ratio Reduction of an OFDM SignalUsing Partial Transmit Sequences", *IEEE Communications Letters*, vol. 4, No.3, pp. 86-88, Mar. 2000.

[5] D. Declercq, G.B. Giannakis, "Recovering clipped OFDM symbols with Bayesian inference", in *Proc. IEEE InternationalConference* on Acoustics, Speech and Signal Processing, pp. 157-160, Jun. 2000.

[6] D.Kim, G.L.Stuber,"Clipping noise mitigation for OFDM bydecision aided reconstruction", *IEEE Communication Letters*, vol.3, 1999.

[7] H.Chen, A.M. Haimovich,"Iterative Estimation and Cancellation of Clipping Noise for OFDM Signals", *IEEE Communications Letters*, vol. 7, pp. 305-307, Jul. 2003.

[8] J.Tellado, L.Hoo, J.M.Cio,"Maximum-Likelihood Detection of Nonlinearly Distorted Multicarrier Symbols by Iterative Decoding", *IEEE Transactions on Communications*, vol. 51, pp. 218-228, Feb. 2003.

[9] M.Deumal, J.Gazda,"Iterative detection of SC-FDMA signals undergoing nonlinear amplification," *Journal on Wireless Communications and Networking*, 2009.

[10] J. Gazda, "Multicarrier based transmission system undergoing nonlinear amplification", Doctoral Thesis, Košice: Technical University of Košice, Faculty of Electrical Engineering and Informatics, 2010.

Keeping smart grid in balance

[1]*Tomáš KOŠICKÝ (3th year)*
Supervisor:[2] *Ľubomír BEŇA*

[1]Dept. of Electric Power Engineering, FEI TU of Košice, Slovak Republic

[1]tomas.kosicky@tuke.sk, [2]lubomir.bena@tuke.sk

***Abstract*—Energy Storage installations have been used in electric utilities for a few decades now. The primary applications of energy storage in utilities include grid stabilization, back-up power and peak shaving, while other potential applications include arbitrage, reduction in renewable variability and frequency regulation. Selection of the appropriate energy storage system for each of the above applications depends on the power/energy ratio requirement, along with the cost, lifetime and other technical challenges. This paper presents a matrix that evaluates the state-of-art energy storage technologies for various utility applications. The characteristics of energy storage systems required for these applications and challenges inherent in the design of such systems are also summarized in this paper. A specific methodology used for analyzing and optimizing the size of the battery energy storage system for frequency regulation is discussed in detail. The same methodology can be extended to the selection of the most optimum technology for other utility applications.**

***Keywords*—Battery storage, Power systems, SVC, STATCOM**

I. Introduction

Energy Storage (ES) installations have been supporting the electric utilities for a few decades now. The primary applications of ES in utilities include grid stabilization, back-up power and peak shaving, while other potential applications include arbitrage, reducing renewable variability and frequency regulation. A key metric to describe an ES characteristic is the Power/Energy (P/E) ratio, which should be matched to the application's requirement. For example, frequency regulation requires high power for a few minutes in either discharge or recharge direction which suggests a high P/E ratio, as opposed to a peak shaving application which requires a comparatively lower P/E ratio and power support time ranging from minutes to an hour or more.

However, with the advent of high P/E ratio ES technologies, for example Lithium ion and commercial high-speed flywheels, they are becoming a viable option for utility applications that feature high P/E ratio use profiles. Improvements in packaging and thermal management solutions for multi-megawatt capacity units have resulted in size and cost reduction for such ES systems. The battery cycle life has also increased to several thousand cycles and typical round trip efficiencies are greater than 85% [1]. Other key reasons for acceptance of ES by utilities are government policies and initiatives on Smart Grids. The intermittent and cyclical nature of many renewable systems puts varying levels of stress on the grid inter-ties, depending on its strength and duration[Fig 1.]. This is especially true for renewable integration on weak/islanded systems where the frequency and voltage variations are more as compared with strongly interconnected grids. The frequency regulation market in UCTE, which is expected to increase in the next decade, is beginning to use such fast responsive energy storage.

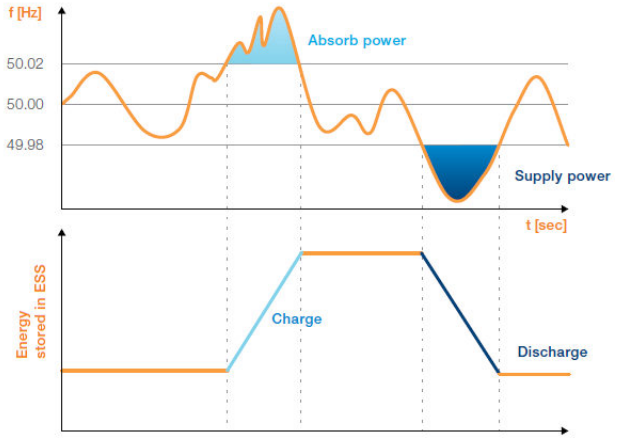


Fig. 1. Frequency regulation with energy storage.

The sizing of ES for such applications involves trade-offs Between initial cost and life cycle cost, especially for **B**attery **E**nergy **S**torage **S**ystems (BESS). There are methods available to predict battery life based on battery **D**epth **o**f **D**ischarge (DoD) and average current rate [2]. Most of these methods do not take the actual duty cycles for the analysis. This paper presents an approach, which uses the real time signals for frequency regulation to predict battery life, and makes trade-offs between the initial capital cost and the life cycle cost. An algorithm to decompose complex cycles to simple DoD cycles is also presented which is used for the battery life prediction based on the data available from battery manufacturers [Fig 2.].

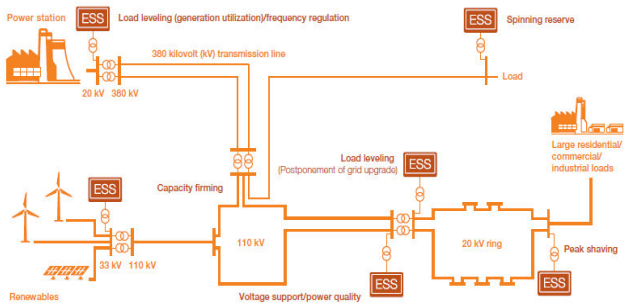


Fig. 2. Typical grid energy storage applications

The sizing of ES for such applications involves trade-offs between initial cost and life cycle cost, especially for Battery Energy Storage Systems (BESS). There are methods available to predict battery life based on battery Depth of Discharge (DoD) and average current rate [3]. Most of these methods do not take the actual duty cycles for the analysis. This paper presents an approach, which uses the real time signals for frequency regulation to predict battery life, and makes trade-offs between the initial capital cost and the life cycle cost. An algorithm to decompose complex cycles to simple DoD cycles is also presented which is used for the battery life prediction based on the data available from battery manufacturers.

Section 2 of this paper discusses the various utility applications of ES, along with the challenges and the energy storage technologies that are the best fit for these applications. Section 3 describes a methodology used forselection of the optimum battery system for frequency regulation. This methodology can be easily extended to other applications.

II. UTILITY APPLICATIONS OF ES

Arbitrage

Arbitrage in financial markets is a practice of taking price advantage due to differences between two or more markets or time periods. Similarly in the electricity trading market, the practice of buying energy when electricity prices are low (Off-peak periods) and selling it when prices are high (Peak periods) is termed as arbitrage. This has been made possible with the development of ES systems, which store the energy and then deliver it later when there is a price advantage. Such an application typically requires a storage time of >2 hrs and a capacity in the range of 10-100s of MW. The technologies used for such an application can have a very low P/E ratio. As large storage capacities are required, Compressed Air Energy Storage (CAES) is a good candidate for arbitrage. During off-peak hours, air is pumped into an underground air-storage "vessel" with electrically driven compressors. This compressed air, at high pressure, is then released through the expander stages during peak hours to drive turbines, which generate electricity. The first commercial CAES system was installed in Hundorf, Germany in 1978 and is rated at 290MW [4]. Another ES technology, which serves the arbitrage market is pumped hydro, where water is stored at an elevated potential during off-peak periods and operates hydro-electric generators during peak hours. More than 90GW of pumped storage is operational worldwide [5]. The advantage of both these systems is long storage duration, high capacities and very long life. However, both of these ES systems have a disadvantage that they have special siting and construction requirements.

BESS are not yet attractive for arbitrage applications due to the relatively lower cycle life and higher normalized energy costs.

Renewable Variability

Wind and solar power present challenges to system operation in that the power output is both variable and uncertain. An ES system provides the flexibility of operating the system in an economic manner, taking into consideration variable tariffs, varying load profiles and intermittent generation.

As an example, wind farms have high power swings due to fluctuation in wind speeds. A closely coupled energy storage can buffer the grid from this variability, e.g. by supplying the power during sudden drops in power output due to loss of wind, as well as help the wind farm meet specified power levels based on the contracts. Moreover, the ES system can help deliver renewable energy at times when it is most needed, i.e. during the peak periods. This applications needs < 2hrs of storage, and the capacity is in a few MWs range.

BESS is the preferred technology, as it meets both the capacity and discharge times required for this application. The BESS can be co-located at the wind or solar farm site itself, and also help improve the power quality by providing "ride-through" for momentary outages. If coupled with advanced power electronics, they can reduce harmonic distortions and eliminate voltage sags and surges. A 34 MW NaS BESS has been used at the 51MW Futamata wind farm in Japan. This system is being used to store the electricity generated during the night [6]. The battery output is combined with the wind turbine-generated electricity during daytime to be supplied to the TEPCO operated power grid.

Ancillary Services

Ancillary Services (A/S) include frequency and voltage regulation, spinning reserves, supplement reserves, replacement reserve and black start capability [7]. The key characteristics of systems used for A/S is that they require very high ramp rates, very quick response time and high reliability. It is expected that the A/S market will grow as the load increases and more renewables (wind and solar) are connected to the grid.

Frequency Regulation

Frequency regulation is required to maintain grid frequency. Independent System Operators (ISOs) procure frequency regulation services in the amount of 1-5% of the daily peak load. These services are generally obtained from generating units that are operated on average at a point lower than their maximum power operating point, giving these units the capability of ramping up or down to address the short-term imbalances in load and generation (MW/minute) [8]. By monitoring the system frequency, corrective action can be taken by these units to minimize the excursion of the power frequency from the expected 50/60 Hz value.

Frequency regulation is served by a variety of power generation sources ranging from pumped hydro facilities and combined cycle plants to steam plants and gas turbines. In

order to provide the least cost of energy, it is necessary to generate as much energy as possible from the lowest cost providers (coal and nuclear). Economic dispatch will dictate that the marginal provider of energy be the unit to provide regulation. Therefore, gas turbines and steam turbines are generally controlled by automatic generation control to respond rapidly to requests for up and down movements and are used to track the minute-by-minute fluctuations in the system load and to correct for naturally occurring power imbalances. Over the last few years, as a result of significant improvements in large-scale, fast-responding Li-Ion batteries and flywheel technologies, alternative frequency regulation sources are now available. As an example, AES and A123/AltairNano are deploying 60MW of BESS containers (2MW, 500kWh for 15 minute storage) [9]. Beacon Power is building a 20MW, 5MWh flywheel system in New York for participating in the frequency regulation market [9]. In addition, ABB and SAFT are pairing up to provide VAR and regulation services with SAFT batteries.

Application	Storage Time	P/E ratio	Ramp rates
Arbitage	4-6 hrs	low	Slow rate recharge at low demand time
Renewable Variability	2 hrs	medium	Moderate ramp rates (MW/30mins)
Frequency Regulation	<30 min	high	High Ramp rate (MW/sec)

Tab. 1. Requirements of various applications

Various governmental directives and regulations are making the entry of storage systems feasible for frequency regulation. A FERC Order 890 ISO/RTO has opened up regulation services markets to storage systems. Storage is defined as a non-generator with less than 1 hour of sustainability, and is regulated around a zero average power base point. Further, in response to FERC Order 890, NY ISO has recently established a Limited Energy Storage Resource market that will enable systems capable of providing energy for 5 minutes or less to now provide frequency regulation in this dedicated market. Three battery systems and one flywheel system will be demonstrating this capability in NY ISO. The MISO Storage Energy Resource allows storage to participate in a one-hour market that is calculated every 5 minutes. ISONE Alternative Technologies Regulation Pilot allows storage to be connected to the system without any bids and is selfscheduled. The payment is based on capacity (Time on Regulation) and mileage (Time on Regulation Service Credit).

III. SELECTION OF OPTIMUM ES SYSTEM

All the applications mentioned above have unique storage requirements, which have been summarized in Table 1. Table 2 lists the technologies suitable for such applications and the efficiencies for such technologies. As is listed, the selection of the storage system will depend on its size, P/E ratio, and also on its Capex and life cycle cost [10]. As most of the applications have dynamic duty cycles, the actual life of the energy storage system, specifically batteries will depend on how it is used in the application. The increased use of BESS for utility applications necessitates the development of a systematic approach to determine the life of the batteries based on the duty cycle. The optimization involves a trade-off between the initial capital cost and life. This approach is discussed in detail in this section.

IV. BATTERY DEGRADATION AND CYCLE LIFE

Batteries are known to degrade in their performance over time. This degradation typically is observed as reduced capacity (where standard charging or discharging protocols yield fewer amp-hours charge in or out of the battery) or increased internal resistance. All batteries have a finite life in any particular application, which is affected by the environment of the battery (particularly thermal) and the way the battery is used electrically. Typically, there is a calendar-life component and a cycling component of the degradation rate. Qualitatively, every charge-discharge cycle results in some degradation. Figure 5 is an example plot of the number of cycles as a function of the subcycle DoD amplitude for an example battery A. For instance, if a battery typically provides 600-700 cycles at 100% DoD, it will provide many more cycles at a lower DoD. It is reasonable to assume that a battery that has a life of X with 100% DoD cycles and proportionally higher life with smaller-depth cycles, would provide more than 2X number of cycles of 50% DoD.

	Flywhell	Battery	Pumped Hydro Storage	Compressed Air Energy
Primary Application	**Power**	**Power/Energy**	**Energy**	**Energy**
Example Application	Frequency Regulation	Frequency Regulation	Arbitrage	Arbitrage
Hr of Storage	1s-15m	15m-1hr	1-10hr	1-10hs
Response Time	10ms	100ms	10s	1min
Efficiency (%)	90-95%	85-90%	70%	50%

Tab. 2. Comparison of various technologies

The effect of DoD on cycle life for some battery systems has been tabulated in the Handbook of batteries [12]. The duty cycle imposed on the battery can be used to estimate the degradation of the ES. The ES system can be sized such that the DoD that it experiences is reduced, though the initial capital cost will increase. The prediction of life based on dynamic cycles is further explored in this paper, with an example of representative duty cycles for frequency regulation.

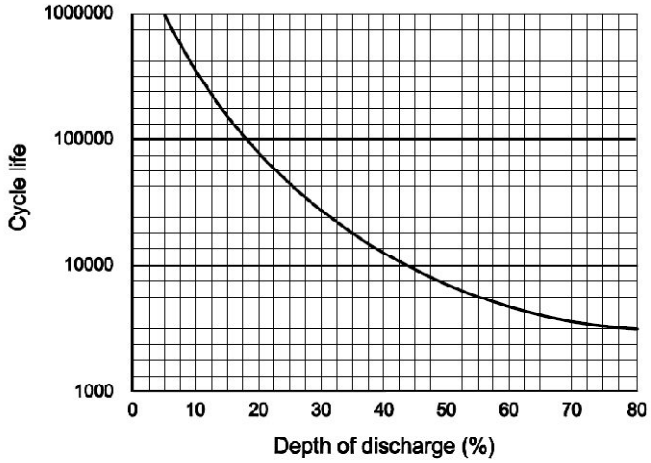


Fig. 3 Cycle-life curve for the Saft lithium-ion battery system

Frequency regulation signals have complex characteristics, based on the difference between generation and demand. Hence, a simple cycle life transfer function cannot be easily applied to ES systems for predicting life. A method of parsing the complex, dynamic cycle into simple cycle events that can be used in aging transfer functions is presented.

V. GENERALIZATION FROM COMPLEX CYCLES TO SIMPLE SUB-CYCLES

To apply the simple cycle aging transfer functions to complex cycles, two fundamental assumptions have been made:

a) complex cycles can be decomposed into a set of equivalent, simple sub-cycles of varying DoD and amplitude, and
b) the aging from the mix of sub-cycles is cumulative, with no interaction between the sub-cycles, and can be summed up from the individual sub-cycle effects.

For analysis, a battery DoD versus time characteristic is decomposed into simple sub-cycles, the aging contribution of each such sub-cycle is determined and all contributions are summed up. As one cycle is defined as a complete charge and discharge from/to the same point, either DoD or State of Charge (SOC) can be used for this analysis. Henceforth, both DoD and SOC have been used interchangeably.
The process of reducing complex cycles into a number of smaller cycles involves what is termed as cycle counting. Rainflow cycle counting algorithm [11] is the most popular algorithm used for analyzing fatigue data in order to reduce a spectrum of varying stress into a set of simple stress reversals. The same algorithm has been used for decomposing the complex DoD cycles into sub-cycles. The flowchart for the algorithm is shown in Figure 4. The S in the algorithm refers to a stack of SOC slope inflection points, where the slope of the SOC changes from positive to negative and vice versa, as a function of time. First, to ensure that half cycles are not counted, the SOC versus time profile is extended to start and end at the same level.

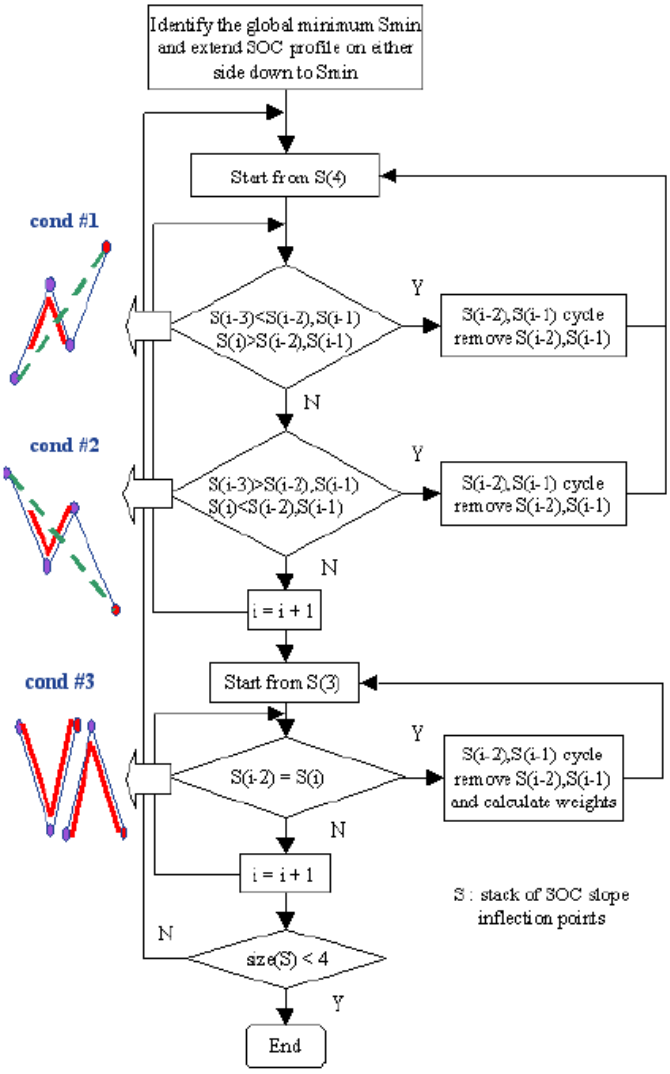


Fig. 4 Flowchart for Rainflow Algorithm

The algorithm starts with the fourth point in the SOC stack, referred to here as S(4), and continues identifying the subcycles based on the three rules as shown in the decision boxes. The three conditions are represented through a simple identified cycle. The algorithm was coded in MATLAB_ to generate the sub-cycle profile. Figure 5 depicts a rather simple duty cycle segment and the identified sub-cycles (three have been highlighted using circles) based on the Rainflow algorithm as described.
Each sub-cycle is assumed to induce a certain degradation in the battery, based on the partial DOD cycle data in Figure 3. When the sub-cycle (DoD) degradation effect is multiplied by the frequency of occurrence of this sub-cycle, it gives the net degradation due to this particular sub-cycle over the entire duty cycle. All such degradations due to the different DoD sub-cycles added would give the net degradation of the battery over the entire duty cycle or profile that it will undergo in such an application.

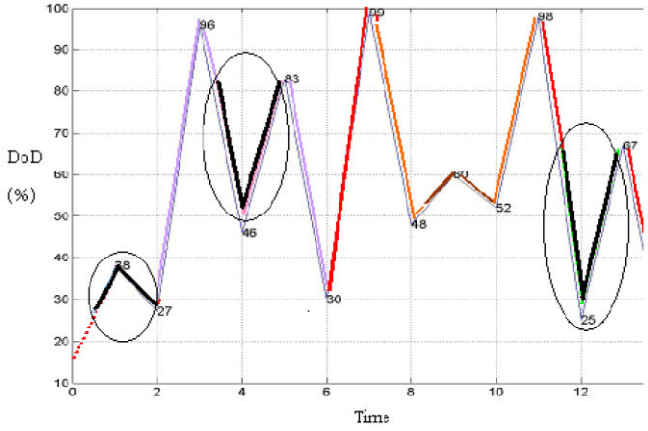


Fig. 5 Identified sub cycles using Rainflow Algorithm

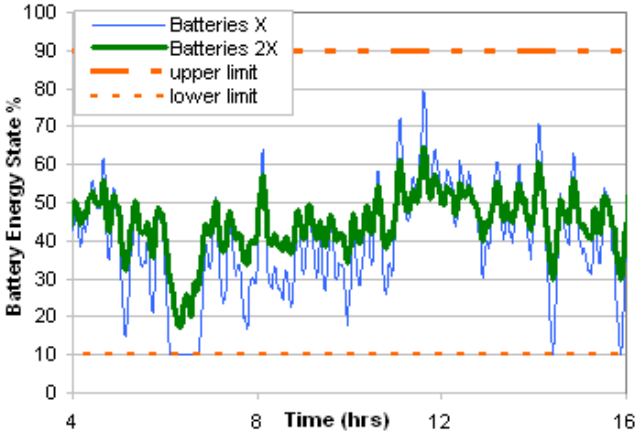


Fig. 7 Battery Energy State estimate vs. time for two battery size

VI. LIFE ESTIMATION OF BESS FOR FREQUENCY REGULATION

A representative frequency regulation signal is shown in Figure 6. The initial sizing of the ES will be done based on the duration of discharge and the power required. For a maximum 5 minute regulation of 1MW, the minimum battery rating results in 1MW, 85kWh. This minimum size would mean that the battery would see relatively large excursions in its charge state.

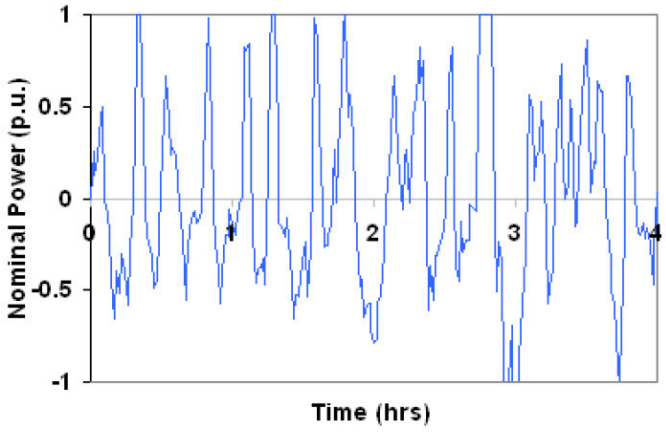


Fig. 6 Example Frequency Regulation Power Trace

As an example, two battery sizes of X and 2X will be chosen to show how the Rainflow algorithm can be applied to analyze the respective battery life. The frequency regulation signal was applied to the battery to find its **S**tate **O**f **E**nergy (SOE) as:

$$\Delta E(t_i) = [MW(t_i) + MW(t_{i-1})] \cdot (t_i - t_{i-1}) / 2$$
$$SOE(t_i) = SOE(t_{i-1}) + [\Delta E(t_i) / Battery\_rating(MWh)]$$

where ti is the instantaneous time in hours and MW(ti) is power required at time ti.
The sign of ΔE(ti) calculated above would tell when the battery is charging or discharging, based on UP/DOWN regulation demands. The 2X battery size experiences lower DoD cycles when compared with the battery size X only. This SOE signal as a function of time is given as an input to the code generated in MATLAB_ to compute the different sub-cycles based on the Rainflow algorithm for the two sizes.

The identified cycles were then used for estimating battery life. It was assumed that the same frequency regulation cycles would be repeated every day. For each DoD subcycle, the center point was calculated (referred to as center SOC) to estimate the number of cycles the battery will last for based on Figure 3. The entire SOC range is divided into several bins (SOC swings), with each represented by a center SOC. As an example, for a DoD of 50-60%, the center SOC will be 0.55. Using the equation from the trend line in Figure 3, the cycle life is estimated as 6063. If there were 4 such 50-60% cycles that the battery would run for every day, the estimated life will be = 6063/(4*365) ~ 4 years.

The SOC bins and associated micro cycles identified by the algorithm are shown in Table 3. However, as the frequency regulation signal comprises of multiple such cycles, the cumulative effect to capacity loss per day is computed, which is then summed up to compute the battery life. The battery **E**nd **o**f **L**ife (EOL) criteria could vary from battery to battery. Some manufacturers consider EOL as the time when the batteries reach 80% of their initial capacity, while others consider lower values. This method of computing degradation to capacity per day due to each cycle helps sum up the total degradation, thereby computing estimated life for 20% or lower/higher capacity loss. The results for the two battery sizes for which the cycles were identified have been tabulated in Table 3. The life has been calculated for two cases where EOL for the battery is capacity lost by 20% in one case and by 40% in the second one. It is clear from the Table that using 2X batteries is a better option under both EOL criteria conditions. Even though the initial cost is double that of the battery size X, the life achieved by using 2X battery size is almost thrice. Hence, from a life cycle perspective, choosing option 2 with double the battery size initially is more economical. These inferences would vary from battery to battery based on the battery life data, as well as the duty cycles that the battery would be subjected to.

Method Advantages and Limitations

A simple method is proposed here which can help optimize the life cycle cost of the batteries used for various utility applications. The method can be extended for calculating life of flywheels as well. However, this algorithm makes some assumptions that one should be aware of:

- It assumes the same regulation signal (cycles) will be repeated every day for degradation computations.
- The results from the life estimation method are based upon the inputs to the algorithm, namely the quality of the battery cycle life plot and the regulation signal.
- The method does not assume that the battery degradation will cause changes in the DoD due to higher battery resistance as the battery is used over life.
- Interactions between sub-cycles to battery life are not assumed.
- Battery operating temperature effects are not accounted for.

VII. CONCLUSSION

The paper highlights the key utility applications, energy storage requirements and challenges for these BES systems. A method was presented for evaluating a battery's cycle degradation under dynamic duty cycles. The method was based on the assumption that the duty cycle is known apriori, and that the battery degradation per micro cycle is independent of macro cycles. The analysis method showed the trade-off between initial capital cost (battery sizing) and the life cycle cost (degradation rate) of the energy storage. A specific example of the methodology is presented for a frequency regulation duty cycle with an example battery.

REFERENCES

[1] Akihiro Bito, "Overview of the Sodium-Sulfur Battery for the IEEE Stationary Battery Committee," by NGK Insulators, Ltd, IEEE PES Annual meeting, 2005.
[2] Study by VRB Power Systems Inc.," The VRB Energy Storage System".
[3] Roger Peters, Lynda O'Malley, "Storing Renewable Power," Primer for Making Renewable Energy a Priority, June 2008.
[4] S.Drouilhet, B.l. Johnson, " A battery Life Prediction Method for Hybrid Power Applications," 35th AIAA AeroSpace Science Meeting and Exhibit, Jan 1997.
[5] Fritz Crotogino, Klaus-Uwe Mohmeyer and Dr. Roland Scharf, "Huntorf CAES: More than 20 Years of Successful Operation," Spring 2001 Meeting, Orlando, Florida, USA, April 2001.
[6] Electricity Storage Association: Large scale electricity storage technologies, http://www.electricitystorage.org/.
[7] James M. Eyer, Joseph J. Iannucci and Garth P. Corey, " Energy Storage Benefits and Market Analysis Handbook," A Study for the DOE Energy Storage Systems Program by Sandia National Laboratories, Dec 2004.
[8] Brendan J. Kirby, "Frequency regulation basics and trends," December 2004, prepared by, OAK RIDGE NATIONAL LABORATORY.
[9] "AES Installs A123 Systems H-APU Energy Storage System at Facility in California, " Battery Power Online, Jan 2009.
[10] Richard Fioravanti, Johan Enslin, "Emissions Comparison for a 20 MW Flywheel-based Frequency Regulation Power Plant", KEMA report, Dec 2006.
[11] Alexandre Oudalov, Tilo Buehler, Daniel Chartouni, "Utility Scale Applications of Energy Storage", IEEE Energy 2030 Atlanta, Georgia, USA, November 2008.

Laboratory Model of a Continuous Processing Line

[1]*Daniel MAGURA (1st year)*
Supervisor: [2]Viliam FEDÁK

[1,2]Dept. of Electrical Engineering and Mechatronics, FEI TU of Košice, Slovak Republic

[1]daniel.magura@tuke.sk, [2]viliam.fedak@tuke.sk

***Abstract*—The contribution deals with design of control algorithms for laboratory model of a continuous production line. The model is equipped by multi-motor drive which are linked by continuously moving strip of processed material. The control system is based on Siemens PLC and FM 458 technological card and the control program is written in the CFC (Continuous Function Chart) language. The operator's panel enables to handle line modes and to input process reference data. The process output data from the line are visualized on a display. The model itself is used in laboratory for training students in multi-motor drive control principles: to clarify processing of continuous strips and to learn hardware and software used for control of the equipment.**

***Keywords*—CFC, continuous production line.**

I. Introduction

The work focuses to control of physical model continuous line consisting of uncoiler, recoiler and three work rolls. Two basic control modes are considered in the model, [1]:
1) tension control of the drives (keeping the tension in the strip among the rolls on a pre-set constant value) and
2) non-tension control where the strip overhang must be kept on a constant value.

The control methods were designed on basis of mathematical model of the electro-mechanical system [2] and were realized using documentation [4] – [7] for the control system T400 from SIEMENS, programmable logic controller Simatic and the SIEMENS card FM 485, in which the control structure in the CFC language is realized.

II. Description of the Continuous Line Physical Model

The physical model (Fig. 1) consists an uncoiler and recoiler and of three work rolls. The arrangement of the work rolls enables to change configuration of the strip line – to use three-, two-, one- or no intermediate roll(s) between the recoiler and unwinder. This enables to develop and use various working modes and control structures of the line drives. There are used DC drives with planetary gear and except of the recoiler they have a worm gear (according to experience the strip tension can be controlled only in the case, when the uncoiler and work rolls are driven though the planetary gear, not through the worm gear).

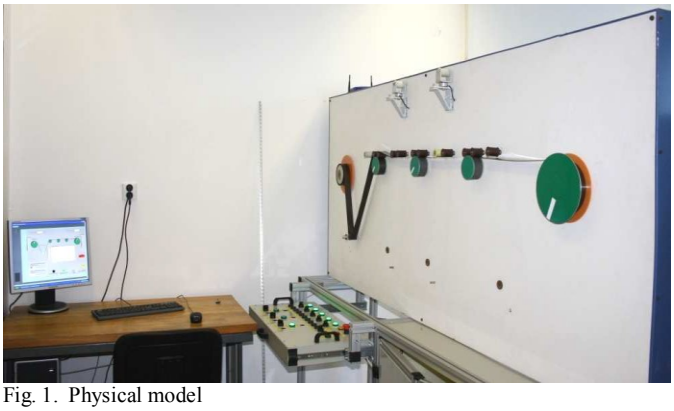


Fig. 1. Physical model

The motors are fed from converters (of the 1386 DC Servo Drive System type). There is a joint DC source with breaking resistor common for all converters. The converters are not equipped by any communication interface and they are controlled by analogous signals generated by technological FM card.

III. Sensors in the Model

The model is equipped by the following sensors – for measurement of the variables:
1) Motor speed - the incremental sensors of the 1XP 8001-1 /1024 type from SIEMENS. Their signals are processed by the technological FM card.
2) Tension in the strip – the strain-gauge sensors of the force of the EMS 30 type are used. The output signal is fed into transducer of the EMS 168 type from the EMSYST Company. The output voltage is proportional to the strip tension. The strain-gauge sensors are capable to measure forces in both directions (pressure and tension). The maximum tension force in the strip is 200 N.
3) Position of the strip overhang. We have used a swinging arm entwisted round a small cylinder (see the arrangement in Fig. 1). The arm angle is measured by one-turn position sensor based on the Hall phenomena (the sensor of 351 HE type from the Vishay company is used). The output voltage of the sensor is within the range from 0,5 V to 4,5 V.

The output of each sensor is fed into the technological card where the signals are processed.

IV. Control of the Multi-motor Drive System

For control of the drives we have used both principles: the tension control and the control without tension in the strip (it is replaced by control of the overhang position). In both cases the model modes are commanded from the control panel and the time courses of variables are visualized on display by the WinCC process visualization program from SIEMENS.

A. Position Control Mode of the Strip Overhang

For a non-tension control three drives in the line are utilized: uncoiler, recoiler and a leading cylinder. The task of control is to keep the strip overhang on a required, preset value. This means, the circumferential speed must be equal on surface of each cylinder. At rewinding the strip from uncoiler, the changing diameter of the uncoiler has to be taken into consideration - by other words, the angular speed of motors depends on the coil diameters as follows:

$$v = \omega.R \Rightarrow R = \frac{v}{\omega}; v = konšt. \tag{1}$$

$$D = D_0 + 2h \cdot N = D_0 + \frac{2.h.\varphi}{2\pi} = D_0 + \frac{2h}{2\pi}.\int_o^t \omega_{ODV}\, dt \tag{2}$$

The position control structure is shown in Fig. 2.

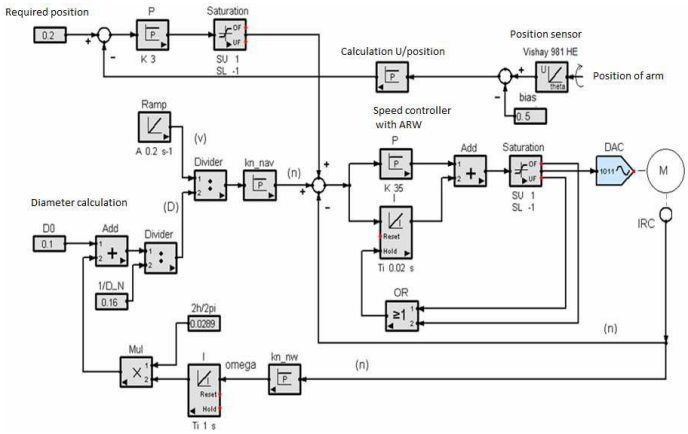


Fig. 2. Control of position at non-tension control mode

We have used a proportional controller with the gain set to K = 3. The controller output is added to the reference signal of the uncoiler revolutions, i.e. any change of the overhang (sensed by angle of the position sensor arm) is ensured by the uncoiler. In this control mode the reference value of the overhang and reference value of the strip is obtained from the visualization programme (WinCC). The recoiler operates in the speed mode, i.e. the speed controller keeps the recoiler motor revolutions on a required value. The uncoiler motor revolutions are influenced by the position controller which tries to align the overhang to its required value. The value of the overhang can be also controlled in the standing mode, at zero speed, of course, when only the uncoiler drive is working. If a larger overhang is required, the necessary length of the strip is unwound from the uncoiler and vice versa.

In Fig. 3 there are shown time courses of reference and real values of the speed, radiuses of the coils, position of the overhang and currents of the uncoiler and recoiler motors. The time courses were measured during the standing and moving mode.

Such solution is used in praxis at a lengthwise cutting of the strip of material into tables (it cannot be done when a tension in the strip occurs). The strip overhang serves also as a small accumulator of the strip when some unforeseeable disturbances occur.

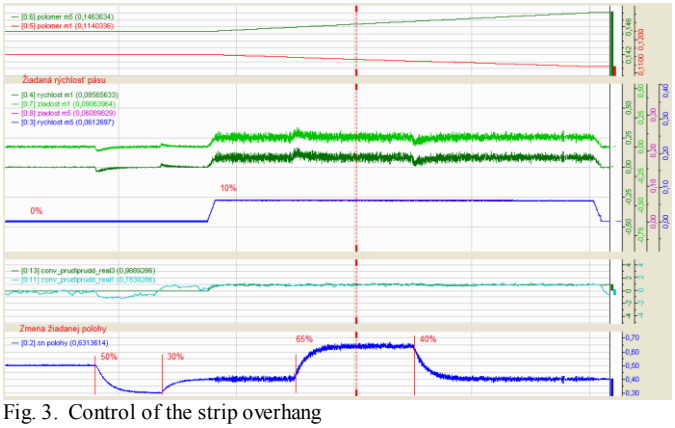


Fig. 3. Control of the strip overhang

B. Tension Control Mode of the Strip

The control structure of the uncoiler drive is shown in Fig. 4. (the control structure of the recoiler is similar to this of the uncoiler in Fig. 2). It is realized using the Siemens T400 system that in praxis is usually used to control drives of uncoiler and recoiler. During the tension control the recoiler winds the strip and the uncoiler works in the generator mode - it is braked and due to this the tension is developed. The principle of tension control consists in knowing the coil radius according to eq. (1) and eq. (2).

$$M_T = F_T \frac{D}{2j_{ODV}} \tag{3}$$

$$\begin{aligned} M_m &= c\phi I \\ M_m &= M_T + M_{comp} \end{aligned} \tag{4}$$

The product of the radius and required tension gives required value of torque on the recoiler shaft, in the control structure, the required torque is derived from the motor current according to eq. (3) and eq. (4). The current is normed and limited by a limiter (what corresponds to the torque limiting). During the operation the condition should be kept saying that the current reference value must by higher than the pre-set limiting.

This condition is fulfilled when the speed reference value is decreased by the value of $\Delta\omega$. In the praxis it means, the uncoiler turns slower than the winder. Fig. 5 shows the process of the tension control during the standstill when the reference value of the tension was stepwise decreased from 35 % to 25 % of the nominal value. A change of the current proportional to the tension is remarkable there.

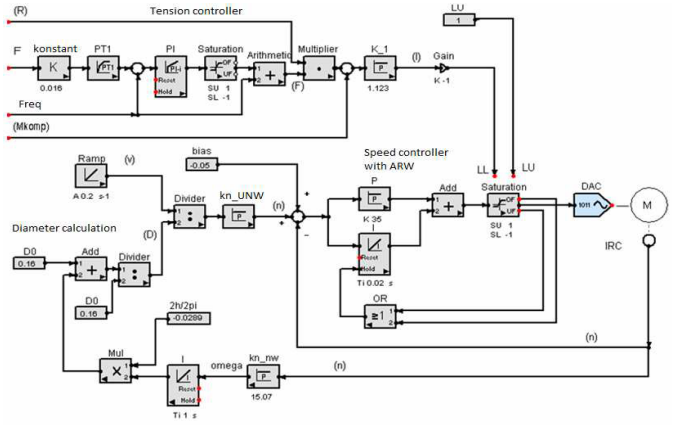


Fig. 4. Tension control

During rewinding the strip the tension controller keeps tension on the reference value with a small error. The largest value of the tension occurs during the drive fall time. We suppose, the deviation is caused by backlash in the worm gear.

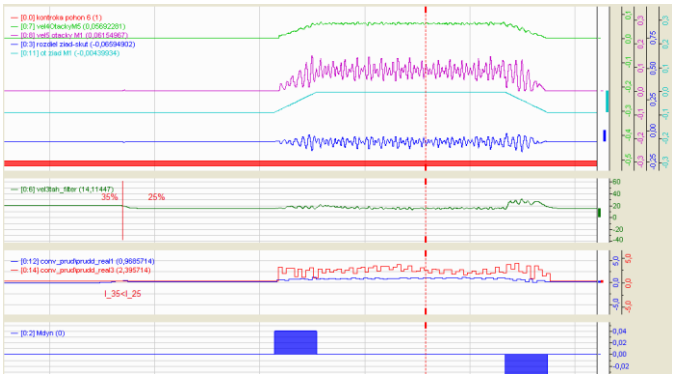


Fig. 5. Control of the tension in web

Fig. 6 shows time responses of the tension control during standstill, from 25% to 65% of the nominal value. In begin we observe tensioning of a released strip (a dashed red line), when the uncoiler turns in a reverse direction by the speed Δω until the time instant while the strip is tensed. After the tensioning there is observable a small overshoot in the tension and then the controller decreases the uncoiler motor current in order to release the tension. When further increasing the tension, the controller decreases the current reference value to the required value.

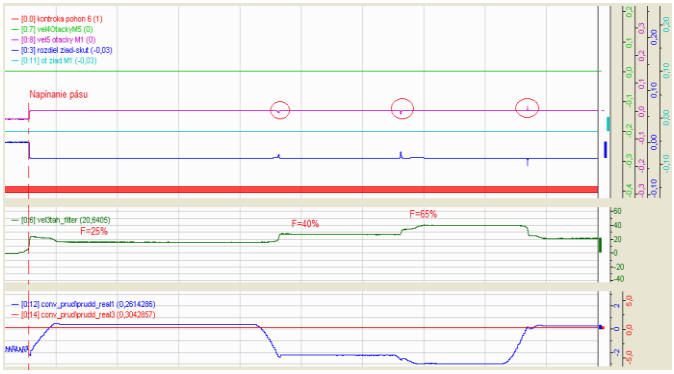


Fig. 6. Tension control during standstill

Because the strip is elastic (it is made from plastic material) we see some overshoots in the time courses of the speed (marked by circles in Fig. 6). Here the uncoiler releases or pulls the strip according to requirement of the tension.

The control structure for tension control has been developed in the CFC language and is completed by working modes, similarly like it is usual in industrial lines: the jogging mode, the mode of tension control and the mode of working speed. The modes are defined in the control program as signals of a Boolean type. The signals are obtained from communication blocks that are interconnected with the program debugged in the LAD language.

V. Visualization of Process

Entire regulating process it is possible to control through the visualization. From visualization it is possible enter process data as analogue attributes like speed, belt overhang or tension. Next it is possible to measure actual process data (time courses) and status like roll diameter, velocity, current, tension and overhang. The visualization was made in WinCC.

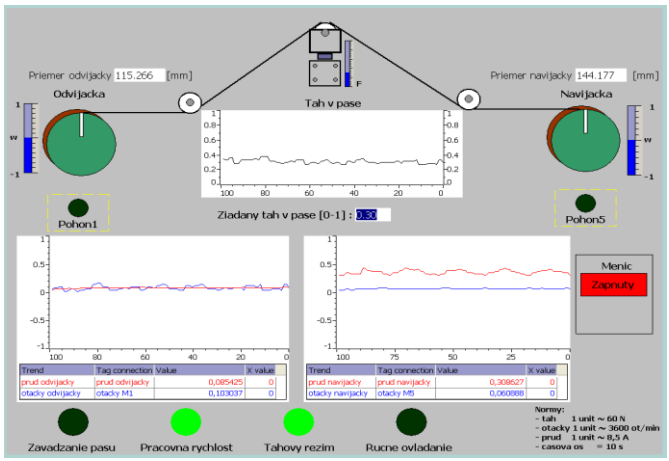


Fig. 6. Visualization at tension control

The visualization for tension control includes time courses about the speed, current and tension, and lamp they showing the actual mode of process (see Fig. 6). The required tension is getting from visualization.

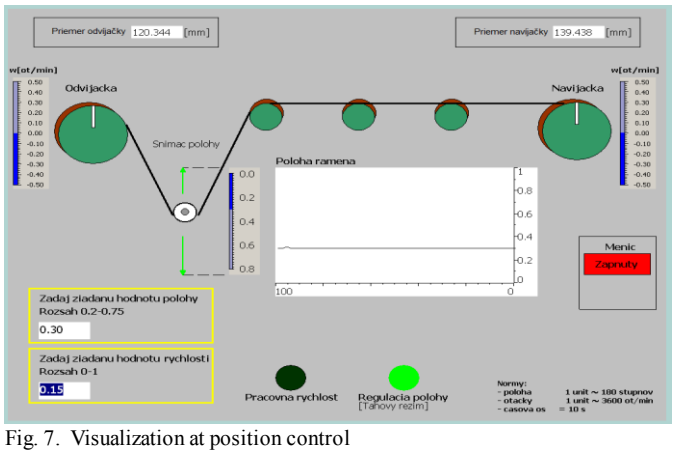


Fig. 7. Visualization at position control

The visualization for position control includes time courses about the position of arm, and lamp they showing the actual mode of process (see Fig. 7). The velocity of belt and required position is possible to getting only from visualization. The mode of process is possible to change only from operator's panel.

VI. Conclusion

Control program is implemented in SIMATIC PLC using the control structures that were programmed in the language CFC. All input and output signals is in the range from zero to one, so we work with the relative magnitudes, it was necessary to standardize the input, and output variables. First, the control structure was simulated in the program MATLAB/Simulink in order to verify the control structures, norms and constants.

In the first we programmed the control structure for direct control of drives. In this case, all drives received the ramp reference speed signal from the control panel. With this structure we also verified functionality of all actuators and sensors used in the model.

In the next step we implemented the non-tension control to a constant overhang. The speed of strip is entered through the visualization panel and is controlled at a constant value. The angular velocities of recoiler and uncoiler are calculated depending of current coil diameter. The diameter is calculated in the technological card with the accuracy lower than half of millimeter.

At the commissioning we used a program *Iba* analyzer to observe signals obtained from the PLC. At tension control, there are used only two drives: recoiler and uncoiler and the tension sensor and in this mode we used empirical setting of the controller parameters that we found easy.

Acknowledgment

The work was supported by Slovak Cultural and Educational Agency of the Ministry of Education of Slovak Republic under the contract KEGA 042TUKE-4/2012 "Teaching Innovation in Control of Mechatronic Systems".

The author thanks to Associate Professor Viliam Fedák, Phd. and Associate Professor František Ďurovský, PhD. for their advices in accomplishment the work.

References

[1] L. Zboray, F. Ďurovský, J. Tomko, *Controlled Drives*, VIENALA 2000, Košice, ISBN 80-88922-13-5, in Slovak.

[2] SIEMENS User Manual, Application Module FM 4851-DP. Document Version A5E01078222-03, Edition 04.2011 Available online: <http://cache.automation.siemens.com/dnl/zg/zg4NDI2MDUA_14952610_HB/FM458-1_DP_e.pdf>.

[3] F. Ďurovský, V. Fedák, "Integrated Mechatronics Systems Laboratory", The 14th Int. Power Electronics and Motion Control Conference, EPE-EPMC 2010, Ohrid, Macedonia. 6-8 September 2010, ISBN 978-1-4244-7854-5.

[4] SIEMENS. Manual, CFC for M7Continuous Function Chart [online]. Document Version A5E00082991-02, Edition 10.2000

[5] M. Hric, Control of Drives of a Continuous Processing Line Physical Model. Diploma thesis, FEI TU Košice, 2009. (in Slovak)

[6] D. Magura, Control of a Continuous Line Physical Model. Diploma thesis, FEI TU Košice, 2012, in Slovak

Lossless image encoding in space of integer discrete wavelet transform

[1]*Ondrej KOVÁČ (2st year)*, [2]*Ján VALISKA (2st year)*,
Supervisor: [3]*Ján MIHALÍK*

[1,2,3] Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

[1]ondrej.kovac@tuke.sk , [2]jan.valiska@tuke.sk , [3]jan.mihalik@tuke.sk

***Abstract* — This paper shortly describes an integer wavelet transformation and its lifting implementation. This paper also describes issue of decomposition of DWT image spectral coefficients represented as integer's numbers to bit planes. Attention is focused on binary state arithmetic encoding. In the conclusion there is a demonstration of encoding using binary state arithmetic code applied on binary planes of DWT representation of images.**

***Keywords* — DWT, Lifting implementation, binary state arithmetic encoding, entropy encoding, image compression, bit planes**

I. INTRODUCTION

The transmission speed of connection has been increasing, but given the fact that the amount of data transferred also increases, it is compulsory to enhance the methods of compression applied to the multimedia data [1][2][3].

II. INTEGER WAVELET TRANSFORM

The biggest disadvantage of DWT is that decomposition coefficients are not integer numbers. In digital signal processing it is better to use integers. In classical DWT convolution of input discrete signal and impulses response of used filters is used. This operation leads to the extension of output discrete signal [2][4]. It is possible to eliminate these disadvantages by using of lifting DWT implementation (LDWT).

A. Lifting DWT implementation

This method is also called fast DWT [4][5][6]. In this approach toward the decomposition of input discrete signal is no extension so it is possible to replace input signal with its DWT representation. Another advantage of LDWT is that inverse DWT can be obtained from direct DWT by simple reshuffle order of operations and signs of summations. With this method, DWT with arbitrary shape of wavelet can be achieved

Let´s suppose a biorthogonal filter bank FB(5,3) [7] which impulses characteristic of filters have odd length and differ by two. Fig.2.1 shows one decomposition level of LDWT.

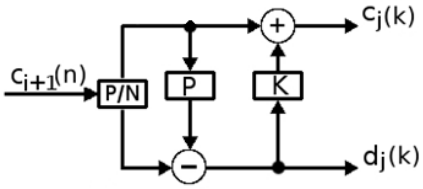


Fig. 2.1 One decomposition level of LDWT

Let´s suppose an input sequence $c_{j+1}(n)$. This sequence is divided into two branches. The bottom branch contains odd samples $b_{j+1}(n)$ and the top branch contains even samples $a_{j+1}(n)$ of input sequence. Essence of this method is prediction of odd samples from even samples through predictor P. Predicted samples are deducted from odd samples and resulting sequence is the detailed coefficients $d_j(n)$ of DWT representation. Coefficients of approximation are results of summation even samples and samples of correction from corrector K. This correction is important, because that sequence of approximation $c_j(n)$ must best reflect the properties of input sequence $c_{j+1}(n)$. Properties of corrector and predictor used in general are dependent on used lifting implementation of filter bank. Output of this decomposition level for FB(5,3) can be described by equations 2.1 and 2.2.

$$d_j(k) = b_{j+1}(n) - \frac{1}{2}\left[a_{j+1}(n) + a_{j+1}(n+1)\right] \quad (2.1)$$

$$c_j(k) = a_{j+1}(n) + \frac{1}{4}\left[d_j(k-1) + d_j(k)\right] \quad (2.2)$$

Obviously, this implementation still does not provide an integer output. With modification of the circuit of fig.2.1 it is possible to get lifting implementation with integer output [5]. Modified circuit representing of integer LDWT (ILDWT) is shown on fig.2.2.

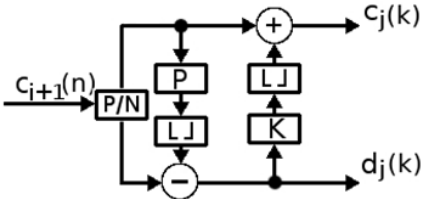


Fig. 2.2 One decomposition level of ILDWT

Equations 2.1 and 2.2 will have this form now:

$$d_j(k) = b_{j+1}(n) - \left\lfloor \frac{1}{2}\left[a_{j+1}(n) + a_{j+1}(n+1)\right] \right\rfloor \quad (2.3)$$

$$c_j(k) = a_{j+1}(n) + \left\lfloor \frac{1}{4}\left[d_j(k-1) + d_j(k)\right] \right\rfloor \quad (2.4)$$

In which operation $\lfloor a \rfloor$ represents rounding to the nearest integer.

Let´s suppose on the input of decomposition level on fig.2.2 input sequence x(n) = [1 2 2 1 0 1 2 3 2 1]. There is a graphical solution of sequence $c_1(n)$ = x(n) decomposition showed on fig.2.3.

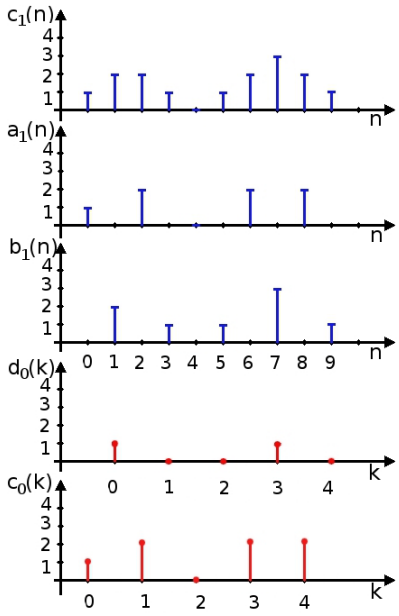


Fig. 2.3 Graphical solution of ILIDWT

Sequences $c_0(k)$a $d_0(k)$ will be rearranged to form $[[c_0(k)]\,[d_0(k)]\,]$ and final integer of DWT representation of input sequence is represented by X(k) = $[[c_0(k)][d_0(k)]]$ = [1202210010]. Fig.2.4 shows one level of inverse ILDWT.

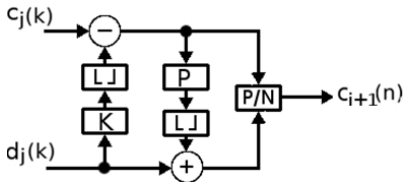


Fig. 2.4 One level of inverse ILDWT

Decomposition to higher order is achieved by engaging of one level to the cascade on the approximation output as showed on fig.2.5.

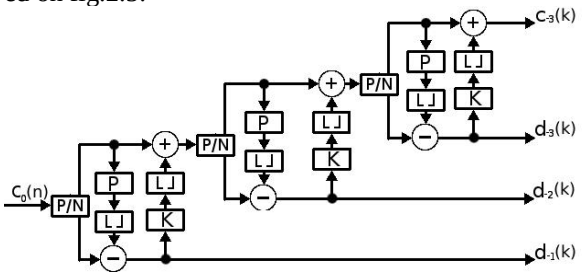


Fig. 2.5 3 levels of ILDWT connected in cascade

2D decomposition is achieved by the application of decomposition into rows and columns. This is shown on fig.2.5 and block schematic is on fig.2.6. For simplification of sketch we suppose that the outputs of P and K are rounded.

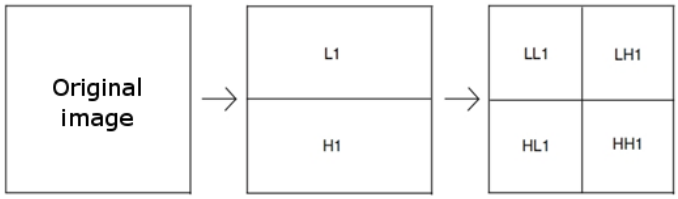


Fig. 2.5 2D ILDWT decomposition

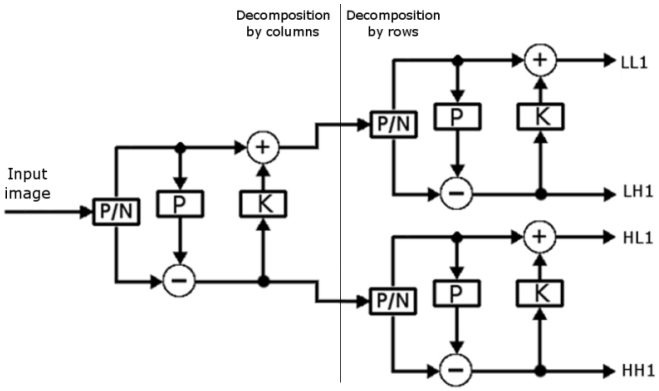


Fig. 2.6 One level of 2D ILDWT decomposition

III. IMAGE DECOMPOSITION TO BIT PLANE

In general, each image can be expressed as two dimension raster of pixels and each pixel can be described by N bits as follows:

$$\text{pixel} = p_0 2^0 + p_1 2^1 + p_2 2^2 + \ldots\ldots + p_{N-1} 2^{N-1}$$

Individual bits P_i correspond to the respective weights 2^i for i=0,1, ... ,N-1. If take only P_0 from in such expressed way pixel from entire polytonal image, we get its bit plane of zero weight (BP_0). By using the same process it is possible create all other bit-planes.

In a transformed space, there are not only spectral coefficients with positive value. Because of this fact, there are two various ways for decomposition.

- *Decomposition of the real bit-plane (RBP)*
- *Decompositions of the absolute bit-plane and plane of sign (ABP)*

A. *Decomposition of the real bit-plane (RBP)*

In this decomposition, we get bit-planes $-BP_{N-1}$ ... $-BP_{0.}$ .... BP_0 ... $BP_{N-1.}$ Fig.3.1. shows the block diagram of RBP decomposer.

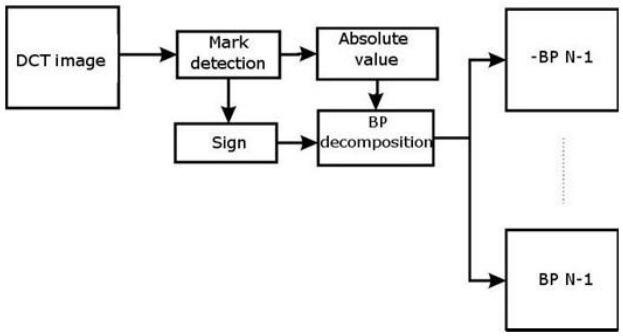


Fig. 3.1 Block diagram of RBP decomposer

Each of spectral coefficients of transformed image is tested on a sign and the result of this is a mark defining either positive or negative bit plane. After this, the image is transformed into absolute value and decomposed with identical method like decomposition of image in Gaussian space.

B. Decompositions of the absolute bit-plane and plane of sign (ABP)

In this decomposition, we get bit-planes BP_0, BP_1 ... BP_{N-1} and BP_s. Block diagram of ABP decomposer is showed on fig.3.2.

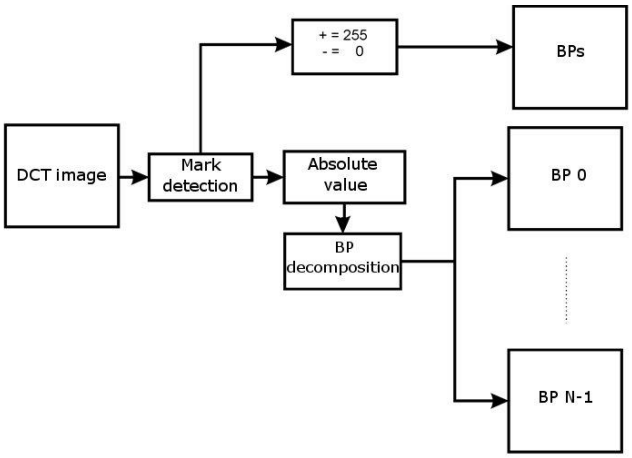


Fig. 3.2 Block diagram of ABP decomposer

Each of spectral coefficients of transformed image is tested on the sign and the result shows value of pixel in bit-plane BP_s. This value is 255, or white for positive and 0, or black for negative mark. After this, the image is transformed into absolute value and decomposed with identical method like in previous case.

IV. ARITHMETIC ENCODING

A. Arithmetic code without using multiplication

1) Encoding

Procedure of the arithmetic [8] encoding (AE) without using multiplications is described with equations 4.1 – 4.4.

$$A(sk) = A(s).2^{-Q} \quad (4.1)$$
$$A(sm) = < A(s) - A(sk) > \quad (4.2)$$
$$C(sm) = C(s) \quad (4.3)$$
$$C(sk) = C(s) + A(sm) \quad (4.4)$$

Where **C(s)** represents binary variable of sequence **S** with 0 initial value before encoding first symbol, and expresses low limit of probabilistic subinterval. **C(sm)** is binary variable of sequence **S** for the more probable symbol and **C(sk)** for less probable symbol of sequence **S**.
Size of probabilistic subinterval is extended by variable **A(s)**, of which initial value is 0.111……11. Variables **A(sk)** and **A(sm)** are binary variables of size of probabilistic subinterval for more and less probable symbol.
This method is based on approximation of probability of less probable symbol by value of 2^{-Q}. Variable **C(s)** is obtained by gradual encoding of symbols of binary variable **S**. When all symbols of binary variable **S** are encoded we get two binary variables, **C(s)** and **A(s)**. Resulted arithmetic code is a binary value from interval <C(s), C(s) + A(s)). This value is made, so it contains the smallest possible amount of valid binary numbers.

2) Decoding

Procedure of decoding is described by equations 4.5 – 4.8.

$$A(sk) = A(s).2^{-Q} \quad (4.5)$$
$$A(sm) = < A(s) - A(sk) > \quad (4.6)$$
$$C(s) < A(sm) => y = m; C(sy) = C(s) \quad (4.7)$$
$$C(s) \geq A(sm) => y = k; C(sy) = C(s) - A(sm) \quad (4.8)$$

Input encoded sequence is decoded with gradual recursion of equations 4.5 – 4.8. Decoder must have information about length of encoded word. This information could be broadcasted in transmission channel separated from encoded sequence, or the length can be given in advance.

B. Binary state arithmetic encoding (BSAC) of binary images

In arithmetic encoding without using multiplications, it is necessary to know value of **Q** from expression 2^{-Q}.
This value must be given for each symbol of the sequence. This information is contained in the model [9]. Model is practically an estimator estimating probability of actual symbol occurrence. JBIG algorithm is used for binary images. Ten points template of JBIG algorithm is on figure 4.1.

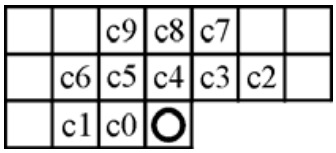


Fig. 4.1. Ten points template of JBIG algorithm.

State **S(y)** of actual point (indicated by circles) is estimated by equation 4.9.

$$S(y) = \sum_{i=0}^{n-1} ci.2^i \quad (4.9)$$

Where **n** is a number of elements in template and **y** is an actual point. After the testing of the whole image, the occurrence frequency of each particular state and its value at each actual point is evaluated. After that, the most frequent occurred value is determined. Finally, conditional probability will be calculated (eq. 4.10.) of point with value "1" or "0" in individual states.

$$p(c = v \mid S_k) = p(S_{v,k}) = \frac{n_{v,k}}{n_k} \quad (4.10)$$

$\mathbf{S_k}$ is value given by equation 4.9 and mean value of actual pixel. $\mathbf{S_{v,k}}$ is state of $\mathbf{S_k}$ with value of pixel equal to **v**.
$\mathbf{n_k}$ is a number of states $\mathbf{S_{v,k}}$. The final model of image is formed by the resulting probabilities. Conditional probability $\mathbf{S_{v,k}}$ are approximated by value 2^{-Q}.

V. EXPERIMENTAL RESULTS

Experimental images CAMERAMAN and BABOON fig.5.1 and 5.2, respectively, were transformed by ILDWT into images of spectral coefficients fig.5.3 and 5.4. ILDWT was used for decomposition of order 3 as it is described in the previous section.

Fig. 5.1 Image Cameraman

Fig. 5.2 Image Baboon

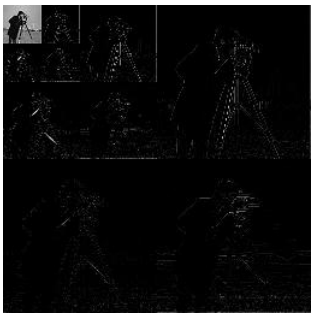


Fig. 5.3 ILDWT image Cameraman

Fig. 5.4 ILDWT image Baboon

Table.1 shows the middle length of code word and reduced redundancy of BSAC applied on the ILDWT images.

TABLE I
MIDDLE LENGTH AND REDUCED REDUNDANCY OF CODE WORD

	Total size bit		Middle length of code word		Reduced redundancy	
Image	Baboon	Cameraman	Baboon	Cameraman	Baboon	Cameraman
IMG	524288	524288	8	8	0	0
RBP	326227	310468	4.978	4.732	3.022	3.268
ABP	286868	269508	5.080	4.815	2.920	3.185

It is clear from table.1 that for image CAMERAMAN which contains more monotonous parts, there is better result achieved for RBP. For image BABOON there is also bigger reduction of redundancy achieved by using RBP decomposition. From reduced redundancy it is possible to express compression using equation 5.1.

$$C = 100\frac{R}{m_o} \qquad (5.1)$$

Where C is compression, R represents redundancy reduced and m_o the middle length of code word of image before encoding. Compression for corresponding cases is showed in the table 2.

TABLE II
COMPRESSION

	Compression	
	Baboon	Cameraman
RBP	37.8	40.9
ABP	36.5	39.8

From table 2 it is clear that, after applying of BSAC on test images in both cases there was achieved a bigger compression for RBP. From table of compression it is clear that compression for image CAMERAMAN is bigger than for image BABOON. This is caused by density of the details. In image with more details there are in ILDWT spectrum more nonzero coefficients of higher frequencies randomly placed in raster. A need of more binary states follows from this.

VI. CONCLUSION

In this paper we described the wavelet transformation with an integer output. For this was used Lifting implementation of DWT based on banks of filters 5,3. Also it was described the modification of ILDWT to the 2D ILDWT. Also it was described the issues of decomposition of image of spectral coefficients to its bit planes. The paper approached the issues of BSAC encoding of LDWT images decomposed onto bit planes without operation of multiplications. In the conclusion the test images were transformed using ILDWT. After that they were decomposed into bit planes and finally encoded using BSAC. From experiment it follows that using of BSAC on to RBP were achieved better results than encoding of ABP.

REFERENCES

[1] MIHALÍK,J.: Kódovanie obrazu vo videokomunikáciach, Mercury-Smékal, Košice, 2001, ISBN 80-89061-47-8.

[2] MIHALÍK, J., GLADIŠOVÁ, I.: Číslicové spracovanie signálov (Návody na cvičenia). LČSOV FEI TU Košice, 1998, ISBN 80-05-00275-0.

[3] MIHALÍK,J.-ŠTEFANIŠIN,R.: Entropické kódovanie v štandardných videokodekoch.Acta Electrotechnica et informatica, Vol.5, No.2, 2005, s.43-50, ISSN 1335-8243.

[4] MIHALÍK, J., ZAVACKÝ, J.: Diskrétne spracovanie signálov, LČSOV FEI TU, Košice, 2012, ISBN 978-80-553-0730-5.

[5] ZAVACKÝ,J.: Celočíselná diskrétna waveletová transformácia. In: Electrical Engineering and Informatics III: Proc. of the Faculty of Electrical Engineering and Informatics of the Technical University of Košice, sept. 2012, p.560-564.ISBN 978-80-553-0890-6.

[6] ZAVACKÝ, J.,MIHALÍK, J., GLADIŠOVÁ, I.: Implementácia diskrétnej waveletovej transformácie v štandarde JPEG-2000. Slaboproudý obzor,roč. 63, č.3-4, 2007, s.5-9, ISSN 0037-668X

[7] ZAVACKÝ, J., MIHALÍK, J.: Banky filtrov pre implementáciu diskrétnej multiwaveletovej transformácie. In: Electrical Engineering and Informatics III: Proc. of the Faculty of Electrical Engineering and Informatics of the Technical University of Košice, sept. 2012, p.554-559. ISBN 978-80-553-0890-6.

[8] GLADIŠOVÁ, I., MIHALÍK, J., ZAVACKÝ, J.: Bezstratová kompresia obrazu pomocou stavového binárneho aritmetického kódovania jeho bitových rovín. Slaboproudý obzor, roč. 62, č.1-2, 2006, s.18-21, ISSN 0037-668X

[9] GLADIŠOVÁ, I., MIHALÍK, J.: Stavové aritmetické kódovanie binárnych obrazov. Acta Electrotechnica et Informatica, Vol.3, No.3, Košice, 2003, p.36-44, ISSN 1335-8243

Magnetoimpedance of Amorphous Ferromagnetic Wire with Small Helical Anisotropy

[1]*Viktória ŠUHAJOVÁ (2nd year)*
Supervisor: [2]Ján ZIMAN

[1]Dept. of Physics, FEI TU of Košice, Slovak Republic
[2]Dept. of Physics, FEI TU of Košice, Slovak Republic

[1]viktoria.suhajova@tuke.sk, [2]jan.ziman@tuke.sk

***Abstract* — Magnetoimpedance measurements on amorphous ferromagnetic $Co_{68.2}Fe_{4.3}Si_{12.5}B_{15}$ wire prepared by in-rotating-water-quenching technique after its thermal treatment with simultaneous application of tensile and torsion stresses are presented. Radial distribution of anisotropy can be expected as a result of this treatment. Qualitative differences were observed on magnetoimpedance loops for low (100 kHz) and high (10 MHz) frequencies of ac current. It was demonstrated that magnetoimpedance interpretation based on skin effect combined with anisotropy distribution can explain the observed behavior.**

***Keywords* — domain wall, impedance, circular magnetization.**

I. INTRODUCTION

Magnetoimpedance effect has been intensively studied during last two decades. The great sensitivity of this effect to very low values of the external magnetic field makes it very promising for technological applications [1].

Magnetoimpedance effect can be observed in magnetic conductors and consists in a significant change of their impedance due to magnetic field. In some materials relative changes in impedance (i.e. magnetoimpedance) can be more than 100 %. This is why this effect is frequently called giant magnetoimpedance (GMI).

The interpretation of this effect is based on the skin effect and for this reason the main parameter for its characterization is so-called skin depth δ. This parameter specifies the thickness of the region beneath the surface of the conductor through which the ac current flows. This means that, if the high frequency current flows through the material, current density is not homogeneously distributed through its cross section but it flows in a certain layer beneath the surface of conductor [2].

Skin depth δ decreases with increasing frequency of ac current f, circumferential magnetic permeability μ_φ and electrical conductivity of the material σ [3]:

$$\delta = \sqrt{\frac{1}{\pi f \sigma \mu_\varphi}} \tag{1}$$

Using the simplest material relationship $\vec{B} = \mu\vec{H}$, where μ is a constant scalar permeability the impedance for cylindrical sample can be expressed as [1, 4]:

$$Z = R + iX = \frac{1}{2} R_{DC} ka \frac{J_0(ka)}{J_1(ka)}, \tag{2}$$

where

$$k = \frac{(1+i)}{\delta}. \tag{3}$$

and J_0 and J_1 are Bessel functions of the first kind, a is the radius of the wire and R_{DC} is dc electrical resistance (for f = 0).

II. EXPERIMENTAL

The magnetoimpedance measurements were carried out on the amorphous ferromagnetic $Co_{68.2}Fe_{4.3}Si_{12.5}B_{15}$ wire with nominal diameter of 125 μm and very low magnetostriction. The sample was prepared using the in-rotating-water-quenching technique. Ferromagnetic wire was treated in two steps – furnace annealing (470 ºC for 3 min) and current annealing (0.5 A for 3 min) with simultaneous application of tensile stress (367 MPa) and torsion of 22 rad/m. The influence if this treatment on the axial hysteresis loops can be seen in Fig. 1 [5].

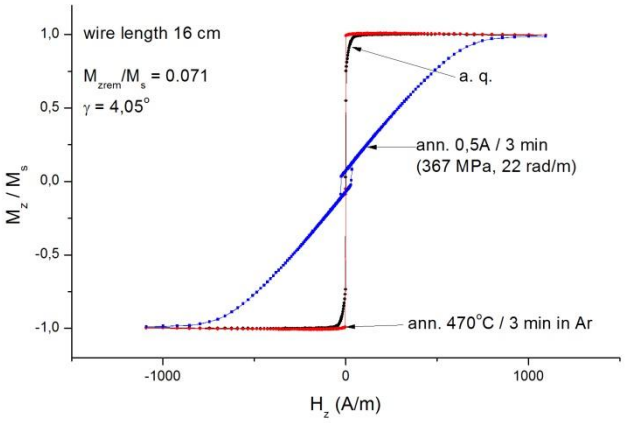


Fig. 1. DC axial magnetization hysteresis loops of the as quenched (a.q.) sample and the same loops after furnace (470 ºC) and current annealing (0.5 A) [5].

The experimental setup for the measurement of circular hysteresis loop presented in this paper is depicted in Fig. 2.

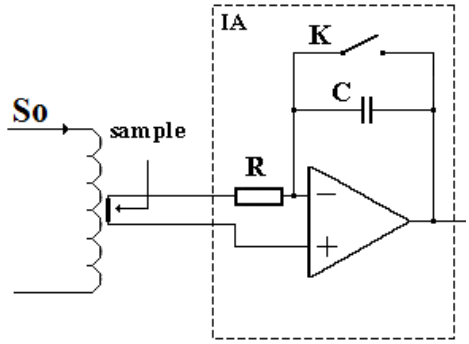


Fig. 2. Experimental setup: So – Solenoid, IA – Integrating Amplifier.

The length of sample used was 1 cm. The thin copper wires were attached to the ends of the sample with silver paint and then connected to the input of an integrating amplifier (IA) which enabled circular magnetization flux to be measured. A homogenous magnetic field was generated in solenoid (So) [6].

The experimental setup used for measurement of impedance is depicted in Fig. 3. Integrating amplifier was replaced by a simple circuit which enabled the ac current to flow through the wire. AC voltage on the sample was measured using high frequency Lock-In amplifier.

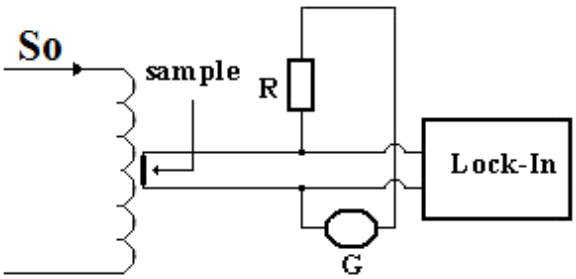


Fig. 3. Experimental setup: So – solenoid, G – power supply of alternating voltage, Lock-In – Lock-In amplifier.

III. RESULTS AND DISCUSSION

Hysteresis loop of circular magnetization flux versus axial magnetic field is shown in Fig. 4. It clearly demonstrates the presence of helical anisotropy. Moreover the shape of the loop in Fig. 4 confirms that the deviation of easy axis from circular direction is small. The magnitude of critical field (H_c) at which circular magnetization reversal takes place is of about 125 A/m.

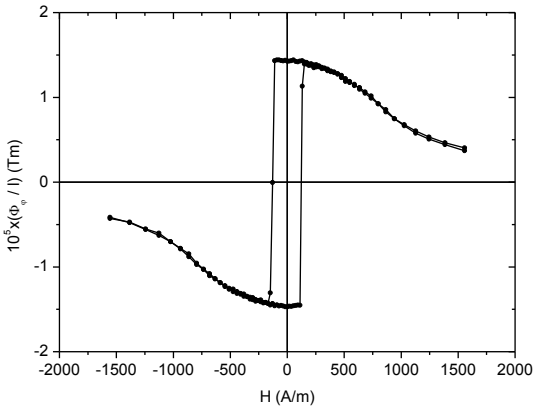


Fig. 4. Hysteresis loop of the circular magnetization flux per unit length of the wire versus axial magnetic field.

The wire impedance for a single point on particular branch of the loop was measured by the following procedure. The impedance was measured in a maximum field (H_{max}) and then the field was changed to its measuring value (H) and the impedance was measured again. Then the field was changed to the opposite value of maximum field (- H_{max}) and finally the field returned to the starting maximum field (H_{max}). The loops of impedance changes $\Delta Z = Z(H) - Z(H_{max})$ measured in this way divided by frequency are plotted in Fig. 5 for two frequencies of ac current flowing through the wire. It can be seen that magnetoimpedance effect is much stronger for high frequency. Besides this fact a few interesting differences can be observed. The so-called two peaks magnetoimpedance curves [7] can be observed for both frequencies. They are typical for samples with transverse magnetic easy axis with respect to the direction of ac current flow and so this is in agreement with the type of anisotropy induced in our sample. These peaks should be observed at anisotropy field. It can be seen in Fig. 5 that the peaks on curves for different frequencies are not observed at the same field. For a higher frequency the magnitude of the field at which the peaks are observed increases. In the low field region the change of impedance for low frequency (100 kHz) is very slow when compared with the curve for high frequency (10 MHz). Also the courses of the loops are different for the two ascending branches of the loops. The branch for the low frequency dependence has a shallow minimum of the impedance in negative field. The circular magnetization reversal (compare Fig. 6a and 6b) causes a decrease of impedance for this frequency. Different behavior can be observed for high frequency. The minimum of impedance is observed in positive field on ascending branch and circular magnetization reversal causes an increase of impedance for this frequency.

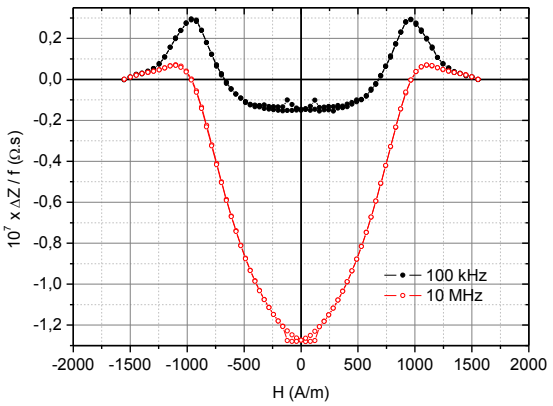


Fig. 5.Hysteresis loops of impedance change divided frequency versus axial magnetic field for two frequencies of ac current.

The interpretation of the behavior described here can be as follows. The sample was thermally treated under simultaneous application of tensile and torsion stresses. While application of tensile stress causes that circular anisotropy is induced in this type of material torsion causes that easy axis direction is at 45° [8] with respect to circular direction. Moreover the magnitude of anisotropy induced by torsion is directly proportional to the distance from the wire axis. Generally it can be concluded that this type of treatment results in radial distribution of the anisotropy. Close to the wire surface a

relative strong helical anisotropy can be expected. Impedance of the wire depends on the skin depth (see Eqs. 1, 2). If we consider the other parameters to be of approximately the same values, the skin depth for the frequency of 10 MHz is ten times lower than for the frequency of 100 kHz. In other words the loop for 10 MHz in Figs. 5, 6 reflects magnetization reversal, mainly circumferential permeability variations along the hysteresis loop, from the surface layer about ten times thinner than the loop for 100 kHz. As was mentioned above close to the wire surface relatively strong helical anisotropy is dominant. The shape of magnetoimpedance loop (for instance minimum of impedance for positive fields on ascending branch of the loop) for 10 MHz in Fig. 5 confirms the presence of this type of anisotropy. Radial anisotropy distribution discussed above combined with much greater skin depth for the frequency of 100 kHz can result in the shift of maximum impedance (peaks position) towards lower fields and in the shift of impedance minimum towards zero field when compared with behavior at high frequency. The last aspect, which this interpretation does not fully explain is the fact that the field at which minimum of impedance is observed, is not only shifted towards the zero field for low frequency of 100 kHz but it even changes the sign. The most probable reason for this behavior is demagnetizing effect. The impedance was measured for the whole relatively short sample so at least close the wire ends demagnetizing effect can be strong.

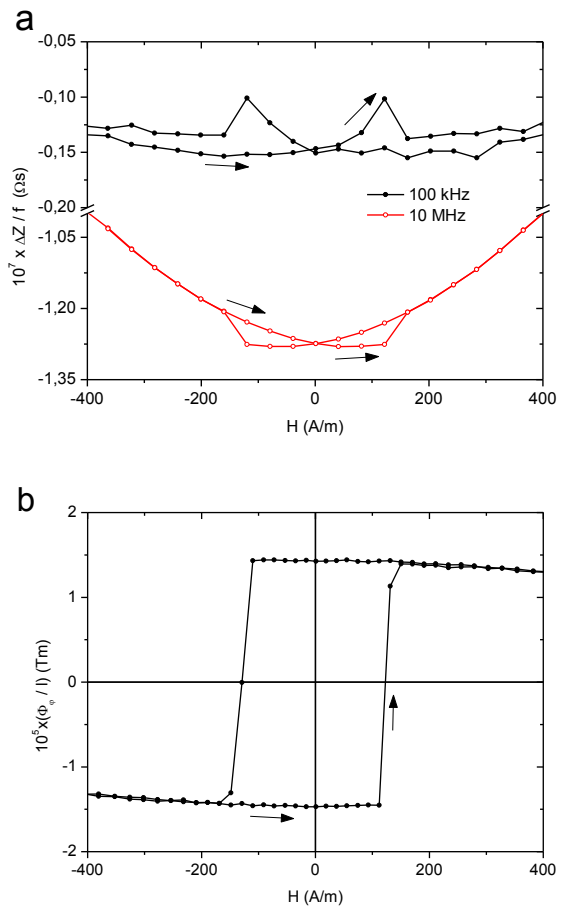


Fig. 6. Low field parts of hysteresis loops of impedance change (a) and circular magnetization flux (b) versus axial magnetic field.

IV. Conclusion

Magnetoimpedance measurements on amorphous ferromagnetic wire after thermal treatment with simultaneous application of tensile and torsion stresses are presented. Radial distribution of anisotropy can be expected as a result of this treatment. Qualitative differences were observed on magnetoimpedance loops for low (100 kHz) and high (10 MHz) frequencies of ac current. It was demonstrated that magnetoimpedance interpretation based on skin effect combined with anisotropy distribution can explain the observed behavior.

Acknowledgment

This work was supported by Slovak Research and Development Agency under the contract No. APVV-0027-11 and also by VEGA grant No. 1/0778/12.

The author would like to acknowledge her supervisor doc. RNDr. Ján Ziman CSc. for proposing the topic of this article and his continuous help.

References

[1] M.Knobel, M. Vázquez, L. Kraus.: *Handbook of Magnetic Materials* (Book style), Elsevier Science D.V.: 2003, pp. 497 - 559.

[2] M. Vázquez.: "Soft magnetic wire," Physics B: Condensed Matter, Vol. 299 (2001), pp. 302 - 313.

[3] D. Jiles.: *Introduction to Magnetism and Magnetic Materials* (Book style), 2nd. ed., Suffolk, 1998, pp. 57 - 60.

[4] Manh-Huong Phan, Hua-Xin Peng.: "Giant magnetoimpedance materials: Fundamentals and applications", Progress in Materials Science, Vol. 53 (2008), pp. 323 – 420.

[5] J. Ziman, J. Onufer, M. Kladivová.: "Domain wall dynamics and Hall effect in eddy current loop in amorphous ferromagnetic wire with small helical anisotropy", Physics B: Condensed Matter, Vol. 406 (2011), pp. 3576 – 3582.

[6] J. Ziman, B. Zagyi.: "DC-.magnetoresistance in surface crystallized FeSiB amorphous wire", Journal of magnetism and Magnetic Materials, Vol. 169, Part 1 (1997), pp. 98 - 104

[7] M. Knobel, K.R. Pirota.: "Giant magnetoimpedance: concepts and recent progress", Journal Of Magnetism and Magnetic Materials, Vol. 242 – 245, Part 1 (2002), pp. 33 – 40.

[8] C. Losin, C. Gómez-Polo, M. Knobel, A. Grishin.: "Torsional Dependence of Second/Harmonic Amplitude of Giant Magnetoimpedance in FeCoSiB Amorphous Wire", IEEE Transaction on magnetic, Vol. 38, No. 5 (2002), pp. 3087 – 3089.

MB2300 - portrait engraver

[1]*Mišél BATMEND (3rd year),* [2]*Marek PÁSTOR (3rd year)*
Supervisor: [3]*Daniela Perduková*

[1,2,3] Dept. of Electrical Engineering and Mechatronics, FEI TU of Košice, Slovak Republic

[1]misel.batmend@tuke.sk, [2]marek.pastor@tuke.sk, [3]daniela.perdukova@tuke.sk

***Abstract*—In this paper, we describe a design and realization of mechatronic system for hard material portrait engraving that we named "mb2300". The paper includes chapters about mechanical design of the machine, electrical design and connection and software/firmware implementation. At the end of paper, engraving results are shown.**

***Keywords*—photo etching, photo engraving, stone processing, portrait engraving**

I. INTRODUCTION

Engraving photos into hard materials such as stone or glass is an old, commonly spread handicraft. Craftsmen have been hand-engraving portraits for years with more or less satisfactory results depending on one's skill. However, even when created by skilled craftsman, it took at least 4 hours of strenuous work to finish a small sized portrait.

Nowadays, a variety of engraving machines is available on a market [1][2][3][4][5][6][7]. The machines are principally computer controlled XY plotters, which position a tool over material, so that it engraves an image based on the input image bitmap. Positioning system typically uses stepper motors in combination with ball screws or toothed belts. The motors and tool are controlled by main controlling unit, which is usually a microcontroller or PLC (programmable logical controller). The main difference among machines is in engraving tool.

Some machines [7] use a laser beam to etch the surface of material. When working with hard materials (e.g. stone, glass), high power of a laser beam is required, which is quite expensive. The advantage is accuracy, which leads to high resolution of engraved image.

Other approach is a small diamond spindle [3][5][6], which scratches the surface by touching the material. The spindle is cheap and engraves fast. Low resolution and low depth of engraving are main disadvantages of this kind of tool.

The third option is a vibrating, sharply pointed diamond [1][2][4]. It engraves by hammering on the material. Hammering technology creates deeper engraving, which widens the range of usable materials. The resolution is also higher than the spindles one. Disadvantage is slower engraving speed.

Concerning any of these tools, result of machine engraving is strongly dependent on quality of the input image. The input bitmap for the machine is always 1-bit depth, black & white (B&W) image. This is because the machine is only capable of creating or not creating intrusion into engraved surface. B&W image is usually created by conversion from tonal photography. Conversion is carried out in personal computer using image processing software (e.g. Adobe Photoshop). Since the conversion is performed in many-to-one fashion, a quality of B&W image is smaller than a quality of original tonal photography. To enhance the quality of B&W image, various image processing techniques must be applied on original image before the conversion is carried out. These techniques include contrast stretching, tonal adjustment, sharpening, noise removal and retouching. As [8] denotes: "The process of enhancing photography is tedious, and requires skill to balance between multiple objectives: contrast in one part of the photography may be traded off for better contrast in another. The craft of photo retouching is elusive and, while a plethora of books describe issues and processes, the decision factors are usually subjective and cannot be directly embedded into algorithmic procedures."

It turns out, that photo enhancement is the key part of engraving process in terms of quality of engraved image. It requires skills which are often underestimated by the machine operators, which leads to unsatisfactory engraving results.

Conversion from enhanced tonal photography into black & white photography can be performed by various algorithms. According to [9], there are many converting algorithms such as: classical screening, direct binary search, error diffusion and other. Depending on the algorithm, the output image gets different look.

II. MB2300 ENGRAVER

As a part of our research, we developed a prototype of engraving machine, which is supposed to engrave portraits into hard materials. We named the machine "mb2300". As an engraving tool, we used the hammering sharply pointed diamond. Development of the machine was primarily based on our preceding model of engraving machine, which we described in [10]. Development could be divided into three main stages: mechanical design, electrical connections design and software/firmware programming.

III. MECHANICAL DESIGN

A. Positioning system

A Frame of the machine was built out of aluminum profiles. Smooth linear movement in all three axes was achieved by using linear bearing guide system. Conversion of a rotary motor movement to linear axis movement was accomplished by toothed belts. We used belts instead of ball screws because they are cheaper and our engraving tool by its principle does not require cutting force. This means, that the motors only have to overcome friction and dynamic force, when moving the axes. As a result, transmitted torque is rather small, so that the belts are stiff enough to keep a desired accuracy of positioning system.

As the actuators in X and Y axis, stepper motors with no position feedback were used. Since the requirements on motor torque are low, we chose the smallest available motors for both axes. In X axis, 42HS03 motor with single ended shaft was used. In bipolar parallel connection its nominal current per phase is 1,4 A, and stall torque 0,47 Nm. In Y axis 57STH76 motor with shaft on both ends was used. In bipolar serial connection its nominal current per phase is 2,0 A, and stall torque is 1,89 Nm. Shaft on both ends was necessary, because there are toothed belts on both ends of a motor in Y axis, which is visible in 3D model in Fig.1. Both motors have angular resolution of 1,8°, which results in 200 steps per revolution. Furthermore, we divide the step by 4 in motor driver, so that we reach 800 steps per revolution. The toothed belts are driven by a 15 teeth pulley with pitch of 2,5 mm. The minimum linear movement of axis is then:

$$k = \frac{15.2{,}5}{200.4} = 0{,}046875\ mm \tag{1}$$

By performing a series of experiments, we found out that optimal minimal linear step is equal to 3.k, which leads to overall resolution of 180,62 dpi:

$$\frac{1inch}{3k} = \frac{25{,}4\,mm}{3.0{,}046875} = 180{,}62\ dpi \tag{2}$$

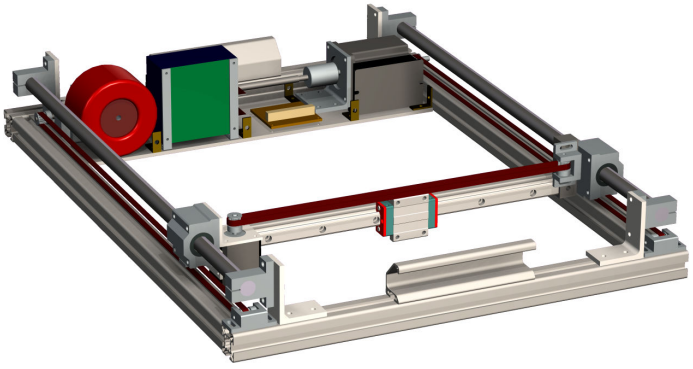


Figure 1. 3D model of machine‘s frame

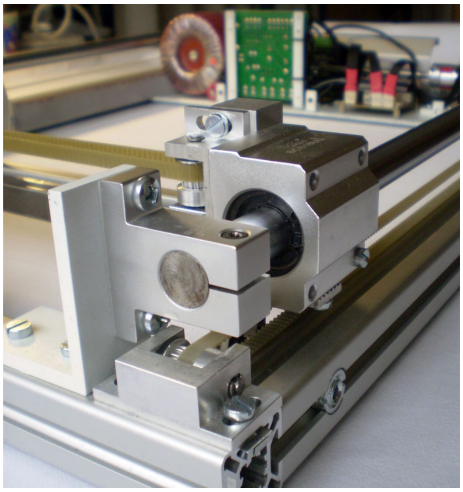


Figure 2. Detailed picture of a built prototype

B. Engraving tool

The tool consists of solenoid and moving plunger ended by a pointed diamond (Fig.3). By energizing solenoid, the diamond hits a solid surface and creates an intrusion. When solenoid is not energized, a spring recovers the plunger to its original position.

Concerning an overall speed of engraving, it is important to examine a maximum frequency of hammering. To explore possibilities of proposed tool, we carried out some experimental measurement on the tool. We attached a position sensor to the tool plunger and plotted the position of the plunger during one stroke (Fig.4).

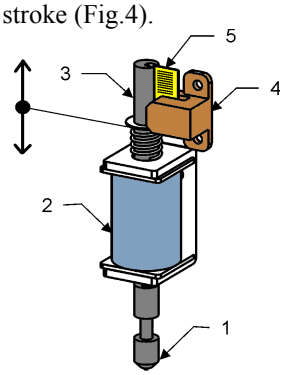


Figure 3. Engraving tool: 1-diamond tip, 2-solenoid, 3-plunger, 4-quadrature encoder, 5-interrupter strip

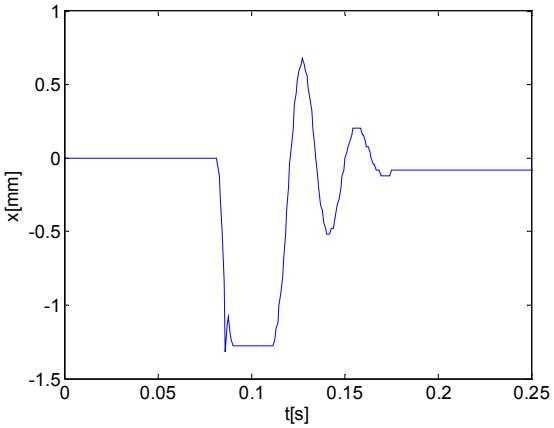


Figure 4. The diamond tip position during one stroke. A full swing of plunger is equal to -2,9 mm. For this application optimal working swing is approximately -1,2 mm. The swing 0 mm refers to a fully opened position when the solenoid is not energized.

The plot was captured by measuring on prototype tool, constructed according to Fig. 3. As a position measuring device, a linear quadrature sensor detecting 25 impulses per mm was used. The time base sampling period was equal to 0,819 ms. The sensor was attached to the plunger core in the way that its influence (additional mass) on measurement was minimized.

As can be seen in a plot, after hitting the desk for the first time, the diamond tip will bounce off and hit the desk again. To disrupt the surface of the desk, energy of the first hit is decisive. Concerning the maximal frequency of strokes, it is sufficient to set the timing, so that the tip will hit the desk only once. The energy of bounce can be then used to help recover the plunger core into zero position (where x=0). With appropriate timing, the energy of overshooting core can be also added to following stroke.

We optimized the control strategy of the tool's solenoid according to previous statements and we performed another measurement which has shown, that the maximum hammering frequency of the tool moves around 58 hits per second.

IV. Electrical Connections

The electrical equipment of the machine includes: power source, control board, motor drivers, motors, electromagnetic tool, tool driver, limit switches, user interface buttons and indicators and wiring.

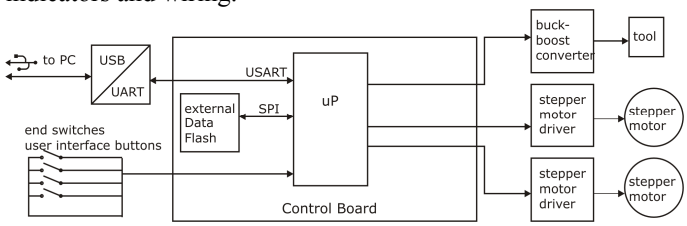


Figure 5. Block schematic of electrical connection

As power source of 108W a toroid transformer and bridge rectifier was used. Its output voltages are 24V/4A and 12V/1A. 24V source is used for power electronics part, 12V source is used for control electronics part.

Both motors are driven by commercial digital driver Leadshine DM422. This is advanced stepper motor driver controlled by microcontroller, which provides many features including microstepping, current loop PID setting, current amplitude setting and other.

As a tool driver, we developed a buck-boost power converter, which controls the solenoid current by voltage PWM. Thus, an electromagnetic tool force which is proportional to current can be controlled. The converter uses HEXFET transistors in both, high and low bridge side. Transistors are driven by IRS2001 integrated circuit. Input impulses for IRS2001 are generated by microcontroller with frequency of 10 kHz. The output voltage of converter is 24 V, maximum current is 2,5 A.

The machine's control system is based on ATmega162 microcontroller from Atmel AVR family. It has 16MHz clock frequency, 16K Bytes Flash, 512 Bytes EEPROM, 1K Byte SRAM memory. It is part of a control board where all necessary peripheral circuitry is implemented. We designed the electrical connection of control board so that it fits the needs of engraving machine. It has sufficient number of inputs and outputs, and uses adequate connectors. Furthermore, it is equipped by 4MB external flash memory AT45DB325D which is used to store the engraving tool path. The communication between microcontroller and memory is carried out by SPI bus.

Based on a tool path stored in a memory the microcontroller controls stepping motor drivers and a tool driver. Two independent timers are used for timing of control pulses for stepping motors. Therefore exact timing for each motor can be achieved.

We built a prototype of control board, which is depicted in Fig.6.

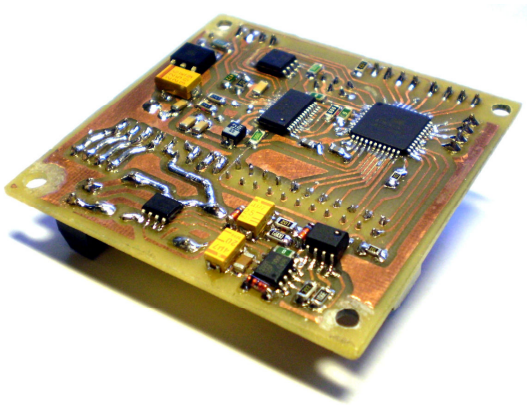


Figure 6. Control board with microcontroller

V. Software and Firmware Equipment

There are three basic program components associated with mb2300 engraver. Firstly, there is an image processing software ran in PC, which is used to adjust and enhance the input image. Secondly, there is a GUI also ran in PC, which allows user to take high-level control over the machine. Finally, there is a microcontroller firmware, which communicates with the GUI and takes a low-level control over the machine.

After a detailed research in available image processing software, we chose Adobe Photoshop Elements as a best program to process images in our application. It offers wide range of fast tools to remove background, adjust tones, sharpen and de-noise image and other. The specific of our application is that one is adjusting grayscale picture in order to make its black & white conversion look good. This is because it is impossible to adjust features like brightness and contrast directly on B&W picture. As a result, the process generally requires following steps: adjust the grayscale image, convert to B&W, evaluate its quality, step back to grayscale and adjust grayscale again and so on. To suppress the number of conversions, we developed a plug-in for Photoshop that allows user to adjust chosen features of grayscale image and directly see what impact it has on its B&W convert. We used Photoshop SDK and its sample filter plug-ins to as a baseline to develop the plug-in. The plug-in is in form of .dll library loaded by Photoshop on runtime. The adjusted image is saved in .bmp format, which can be loaded by machines GUI.

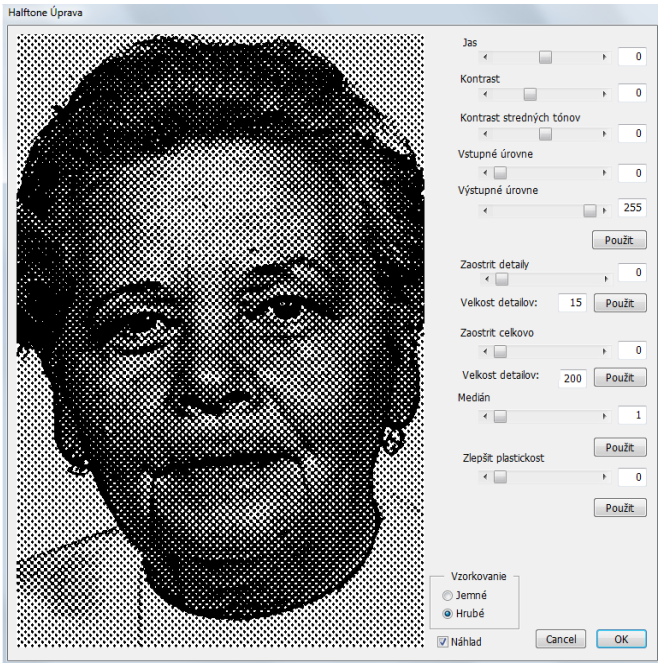


Figure 7. Halftone adjust Photoshop plugin

The GUI of the machine was developed using C# programming language. Its main aim is to allow user to load B&W image, to adjust size and rotate image, to set the starting position for engraving and to start/pause engraving. The GUI communicates with machines control system through USB/serial convertor.

The microcontroller's firmware was programmed using C programming language. It has over 1700 program lines and includes routines for serial communication with PC, SPI communication with external flash, generating motor impulses, generating tool driver impulses, handling limit switches and user buttons, engraving based on tool path stored in external memory, handling unexpected situations (e.g. power loss) and other.

The process of engraving looks as follows:

1. The user adjusts the image in Photoshop and saves it as B&W bitmap.
2. The user opens adjusted bitmap in GUI, sets desired dimensions and starting point of engraving. The bitmap is then converted to tool path and the tool path is fetched into microcontroller's external memory.
3. Once the starting command is received, microcontroller reads the tool path data action by action and controls stepping motor drivers and a tool driver accordingly.

VI. Results

Using a built prototype, we performed many experiments and engraved over 400 portraits into polished granite stone and a few portraits into glass. Engraving results reach quality which is comparable to other engraving machines. Moreover, with a proper control of the tool hitting force, we are able to engrave into non-black granite with extraordinary results. The complete machine with aluminum cover is depicted in Fig.8. The examples of engraved portraits are shown in Fig.9.

Figure 8. mb2300 engraving machine

Figure 9. Examples of engraved portraits

References

[1] Photomaster - stone image engraving machine , Abra, PL, available online: www.abra.pl/en

[2] PhotoGrab - stone image engraving machine Helios,IT, available online: www.heliosautomazioni.com

[3] N-Graver - stone image engraving machine, Schubert Software, DE, available online: www.schubert-software.de

[4] Photo Marker - stone image engraving machine, Cobalm, IT, available online: www.cobalm.com

[5] G-Tech - stone image engraving machine, Boris Malacko, DE, available online: www.itys.sk

[6] G10- stone image engraving machine, Galeks, Acmetek, SRB, available online: www.acmetek.biz

[7] Z12 Monument Etching Laser, available online: http://www.kernlasers.com/product_pages/z12.htm

[8] BYCHKOVSKY, V. – PARIS, S. – CHAN, E. – DURAND, F.: Learning photographic global tonal adjustment with a database of input / output image pairs, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 97-104, ISBN: 978-1-4577-0394-2, June 2011

[9] V. MONGA, N. DAMERA-VENKATA, B. L. EVANS,: Halftoning toolbox for MATLAB. The University of Texas at Austin, available online: http://users.ece.utexas.edu/~bevans/projects/halftoning/toolbox/

[10] BATMEND, M.: Fyzikálny model polohovacieho stroja, bachelor thesis, Košice: Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky, 2008, 69 pages.

Measuring of dependence of shielding effectiveness of electromagnetic field on the distance in high frequency range

[1]*Marek PAVLÍK (2st year),* [2]*Ján ZBOJOVSKÝ (2st year)*
Supervisor: [3]*Iraida KOLCUNOVÁ*

[1,2,3]Dept. of Electric Power Engineering, FEI TU of Košice, Slovak Republic

[1]marek.pavlik@tuke.sk, [2]jan.zbojovsky@tuke.sk, [3]iraida.kolcunova@tuke.sk

***Abstract*— This article deals with measurement of dependence of shielding effectiveness on the distance in high frequency range. The measured object is Medium Density Fiberboard wood board with thickness 12mm (MDF12).**

***Keywords*—shielding effectiveness, MDF12, electromagnetic field.**

I. INTRODUCTION

Currently, we are witnessing an increase in resources, electromagnetic radiation, which in the past was not so great. Wireless on the one hand, provides convenience and speed of communication, but on the other hand brings some negatives that are in the world, more and more discussed topic. It should be noted that the sources of the electromagnetic field not only negatively affect the human body. A well-known example of the beneficial uses of electromagnetic radiation source at higher frequencies, the thermal hytertermia, used in the treatment of tissue, which is described in [1].

Wireless communication is not only represented in the IT sector, but also the constant development in the medical environment requires the use of this type of communication. In the medical field, there are many applications using wireless communication just for capturing and monitoring of the human body, the output data are collected as the human body.

The wireless communication provides long-term monitoring activities of the human body even under severe conditions [2].

Electromagnetic radiation we are exposed to every time and everywhere, and most likely, the trend of increasing sources of electromagnetic radiation will continue to increase. As an ordinary person can not with our senses perceive electromagnetic radiation is so benevolent to potential threats. Precisely because it is a topic of increasing sources of electromagnetic radiation and its impact on actual human for the general public.

Stand opposite each other, on the one hand, manufacturers of devices operating on the principle of electromagnetic field, claiming that their funds are safe and on the other stands the World Health Organization which say opposite. Since the period of use of such devices is relatively short, we can not say with certainty whether these devices cause health risks or not.

Wide professional community more and more focused to resources of electromagnetic fields, their impact and the associated concept of electromagnetic compatibility.

II. SHIELDING OF ELECTROMAGNETIC FIELD

Wide professional community more and more resources devoted to electromagnetic fields, their impact and the associated concept of electromagnetic compatibility. EMC could be defined as the ability of devices to coexist in the same electromagnetic environment. Electromagnetic interference can cause severe problems, and it should be taken into account in the design of new power plants. [3] The term Electromagnetic Compatibility is closely related to the notion of shielding electromagnetic field.

Given that any material shade, respectively. prevents the penetration of the electromagnetic field, in this article is focused on shielding effectiveness of Medium Density Fiberboard with thickness 12 mm.

Shielding use different types of equipment in various fields of our daily life, from mobile phones, mobile stations, wifi devices to the Internet through a variety of medical devices, communications networks, electronic devices and the like. [4] [5].

Quality shielding material are determined by three coefficients, K_S shielding coefficient, absorption coefficient and a reflection coefficient R. With these three factors closely related to the shielding effectiveness *SE*. Shielding coefficient K_S is determined by the electric field strength E, possibly based on the intensity of the magnetic field H by the relation:

$$K_S = \frac{E_2}{E_1}; K_S = \frac{H_2}{H_1} \quad (1)$$

where:

E_2- electric field measured using the antenna placed in the prescribed configuration within the enclosure

E_1- electric field measured using the antenna placed in the prescribed configuration in the absence of the eclosure

H_2- magnetic field measured using the antenna placed in the prescribed configuration within the enclosure

H_1- magnetic field measured using the antenna placed in the prescribed configuration in the absence of the eclosure

Shielding effectiveness *SE* is calculated using the formula:

$$SE = 20.\log\frac{1}{|K_S|} = 20.\log\frac{|H_1|}{|H_2|} = 20.\log\frac{|E_1|}{|E_2|}[dB] \quad (2)$$

Formula for determining the effectiveness of shielding *SE* according to the frequency range changes. According to [5], the screening efficiency is determined by the relationship:

$$SE = 20.\log\frac{|H_1|}{|H_2|} = 20.\log\frac{|V_1|}{|V_2|}[dB] \quad (3)$$

For the frequency range 50 Hz to 20 MHz, where:

H_2- magnetic field measured using the antenna placed in the prescribed configuration within the enclosure

H_1- magnetic field measured using the antenna placed in the prescribed configuration in the absence of the enclosure

V_2- voltage reading within the enclosure

V_1- voltage reading in the absence of the enclosure

$$SE = 20.\log\frac{|E_1|}{|E_2|} = 10.\log\frac{P_1}{P_2}[dB] \quad (4)$$

For the frequency range 20MHz to 300MHz and also the same applies to the frequency range 300MHz to 100GHz where:

E_2- electric field measured using the antenna placed in the prescribed configuration within the enclosure

E_1- electric field measured using the antenna placed in the prescribed configuration in the absence of the enclosure

P_2- power detested within the enclosure

P_1- power detected in absence of the enclosure

According to [6] is the sum of the electromagnetic field shielding reflection *R*, multiple reflection *B* and absorption *A* of electromagnetic field derived as:

$$SE = A + R + B$$

$$SE = 15{,}4t\sqrt{\mu\sigma} + 168{,}16 - 10\log\frac{\mu_R.f}{\sigma_R} + 20\log\left(1 - e^{\frac{2t}{\delta}}\right) \quad (5)$$

where:

t - material thickness

σ - conductivity of shielding material

μ_R - permeability of shielding material

f - frequency

δ - depth of penetration

For simplicity, it is possible to determine the shielding effectiveness *SE* as:

$$SE = A + R$$

$$SE = 8{,}69.\frac{t}{\sqrt{\frac{2}{\omega.\mu.\sigma}}} + 20.\log\left(\frac{1}{4}.\sqrt{\frac{\sigma}{\omega.\mu_r.\varepsilon_0}}\right) \quad (6)$$

kde

t - material thickness

δ - depth of penetration

μ - permeability which is also included permeability shielding material

σ - conductivity of shielding material

μ_R - permeability of shielding material

ε_0 - permittivity of vacuum

From equation (6) shows that the shielding effectiveness *SE* affects relative permeability and relative conductivity [6].

III. WORKPLACE FOR THE PURPOSE OF MEASURING SHIELDING EFFECTIVENESS OF ELECTROMAGNETIC FIELD

Connection workplace for the purpose of measuring shielding effectiveness of electromagnetic field is shown in Fig. 1.

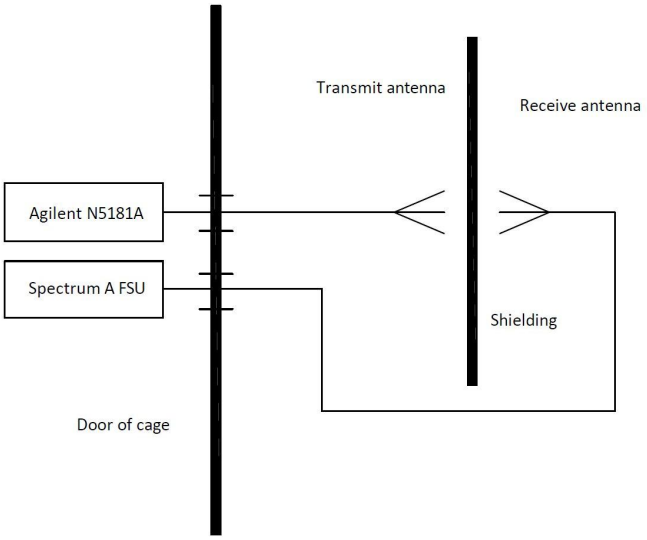


Fig. 1 Workplace for measuring shielding effectiveness electromagnetic field

For the purpose of measuring the effectiveness shielding used pulse generator Agilent N5181A. The pulse generator was adjusted value of 17.88 dBm. Before the measurement was made, calibration cable and the associated calibration antennas was made. It is for this reason, because we set the value of 17.88 dBm. Cable attenuation from pulse generator to transmitting antenna was 1.88 dBm. It follows that the power from transmit antenna to receive antenna is 16 dBm.

Transmitter and receiver antenna are the type of Horn. For a complete calibration workplace the calibration curves were used. Cables were calibrated according to the calibration curves antennas.

IV. RESULTS

Figure 2-5 we can see the dependence of value of an electromagnetic field on the distance from the transmitting antenna of the frequency of 1 GHz (Fig. 2), 2 GHz (Fig. 3), 3 GHz (Fig. 4) and 4 GHz (Fig. 5).

The graph shows that the higher frequency wooden plate shade more than the lower frequencies. For example, at a distance of 100 cm was shielding effectiveness as follows:
At a frequency of 1 GHz, *SE* = 0.1 dB
At a frequency of 2 GHz, *SE* = 1.1 dB
At a frequency of 3 GHz, *SE* = 1.8 dB
At a frequency of 4 GHz, *SE* = 2.6 dB

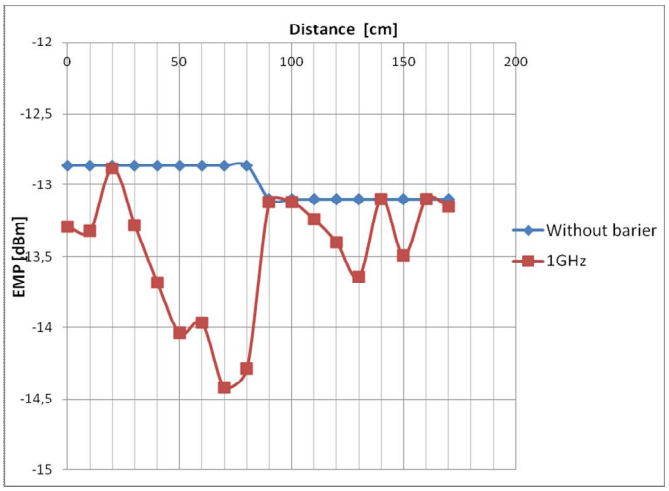


Fig. 2 The Dependence electromagnetic field on the distance for frequency of 1 GHz

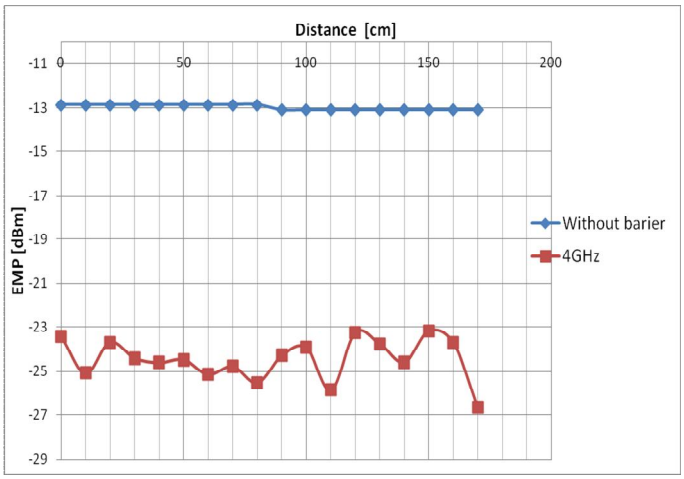


Fig. 5 The Dependence electromagnetic field on the distance for frequency of 4 GHz

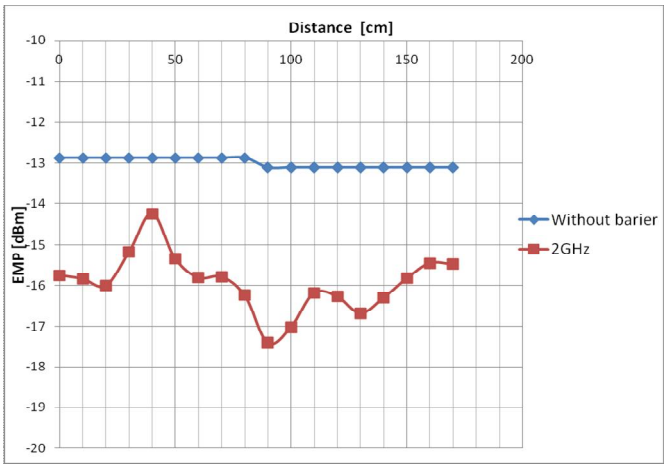


Fig. 3 The Dependence electromagnetic field on the distance for frequency of 2 GHz

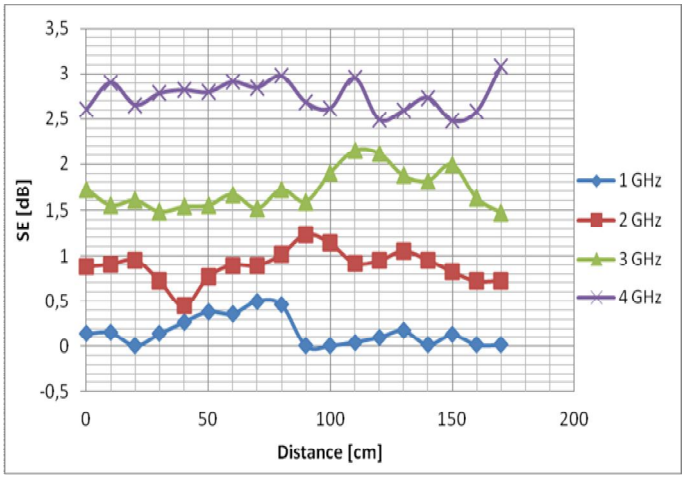


Fig. 6 Shielding effectiveness for frequency 1 GHz, 2 GHz, 3 GHz and 4 GHz

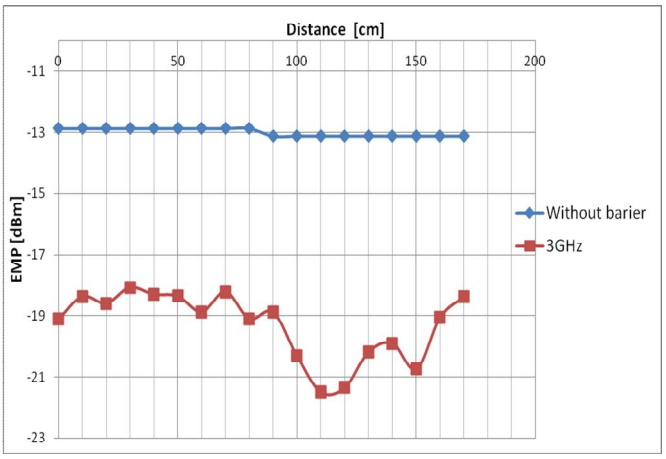


Fig. 4 The Dependence electromagnetic field on the distance for frequency of 3 GHz

V. Conclusion

This article was foused at measuring the effectiveness of shielding MDF12 wooden board. Thickness of the wooden plate is 12 mm. Measurements were made at frequencies of 1 GHz, 2 GHz, 3 GHz and 4 GHz. She was also a wooden board moved from the transmitting antenna in the range from 0 cm up to 170 cm.

Need to know the material as a shadow is currently considerable. The most basic parameter is the shielding effectiveness of the shield. For sure, true that metal materials have better shielding effect than other building materials such as wooden board, which has been measured by the measurement object.

Acknowledgment

This publication is the result of the Project implementation: Protection of the population of the Slovak Republic from the effects of electromagnetic fields, ITMS: 26220220145, Activity 2.1 supported by the Research & Development Operational Programme funded by the ERDF .

We support research activities in Slovakia / Project is cofinanced from EU funds.

REFERENCES

[1] Habash R.W., Krewski D., Bansal R., Alhafid H.T., Principles, Aplications, Risk and Benefits of Therapeutic Hyperthermia, Frontiers in Bioscience (Elite Edition), 3 (2011), 1169-1181.

[2] Plawiak-Mowna A., Krawczyk A., Wireless Body Sensor Network-Fundamental Concepts and Apalication, Przeglad Electrotechniczny, 12b/2012, 2012, s.267-268.

[3] Kmec M., Hvizdoš M., Skúšky digitálnych ochrán prístrojom Omicron CMC, Electrical Engineering and Informatics 3 : proceeding of the Faculty of Electrical Engineering and Informatics of the Technical University of Košice. - Košice : FEI TU, 2012 S. 705-710. - ISBN 978-80-553-0890-6

[4] Rusiecki A – Calculation and measurements of shielding effectiveness of slotted enclosure with built-in conductive stirrer, Przegland Elektrotechniczny, ISSN 0033-2097, R. 88 NR 10b/2012, s. 328-329.

[5] IEEE Standart, Method for Measuring the Effectiveness of Electromagnetic Shielding Enclosures, EMC Society, New York 2006, p.39.

[6] Pan Z., Zhang H., et al.: Advances of Studies on Electromagnetism Shielding Fabric, Science & Technology Review, 27(2009), 24, 86-91.

Measuring of shielding characteristics of shielding chamber

[1]*Marek PAVLÍK (2st year),* [2]*Lukáš LISOŇ (1st year)*
Supervisor: [3]*Iraida KOLCUNOVÁ*

[1,2,3]Dept. of Electric Power Engineering, FEI TU of Košice, Slovak Republic

[1]marek.pavlik@tuke.sk, [2]lukas.lison@tuke.sk, [3]iraida.kolcunova@tuke.sk

***Abstract*— This article deals with measurement of shielding chamber against penetration of electromagnetic field into the chamber. The measurement is focused on the frequency range at which mobile phones operate.**

***Keywords*—chamber, electromagnetic field, shielding effectiveness.**

I. INTRODUCTION

Mobile Communications is currently very widespread. Because of the large spread of mobile communication is becoming more and more interested for the general public about their impact on the human body.

This paper is focused on determining the characteristics of properties electromagnetic field inside the shielding chamber. This article is focused on shielding electromagnetic field at frequencies on which mobile phones work.

The frequencies used for mobile networks can be divided into four zones:

1. *Primary* (900 MHz) - while the band is divided into the upload band (890-915 MHz) band and for download (935 – 960 MHz)
2. *Second* Advanced (900 MHz) – to 880-915 MHz upload, download for (925-960 MHz)
3. *DCS1800* (1800 MHz) – for 1710-1785MHz upload, download for 1805-1880 MHz
4. *DCS1900*

Another example of a rapidly expanding electromagnetic field of high frequency is wi-fi communications.

Wireless communication is not only represented in the IT sector, but also in the medical environment requires the use of this type of communication. In the medical field, there are many applications using wireless communication just for capturing and monitoring of the human body, the output data are collected as the human body.

Wide professional community more and more focused to resources of electromagnetic fields, their impact and the associated of electromagnetic compatibility.

This paper is focused on determining the characteristics of shielding chamber. Specifically focused on determine the characteristics of against the effects of electromagnetic fields. Variable determining the shielding of electromagnetic field is shielding effectiveness *SE*. The shielding effect is closely related to the concept of electromagnetic compatibility. Electromagnetic compatibility could be defined as the ability of devices to coexist in the same electromagnetic environment. Electromagnetic interference can cause severe problems, and it should be taken into account in the design of new power plants. [3]

II. MATHEMATICAL MODEL

Mathematical model of the shielding effect can be determined using the formula:

According to [2] shielding effectiveness is derived as:

$$SE = A + R + B$$

$$SE = 15{,}4t\sqrt{\mu\sigma} + 168{,}16 - 10\log\frac{\mu_R . f}{\sigma_R} + 20\log\left(1 - e^{\frac{2t}{\delta}}\right) \quad (1)$$

where:

- R - reflection coefficient
- B - multiple reflection coefficient
- A - absorption coefficient
- t - material thickness
- σ - conductivity of shielding material
- μ_R - permeability of shielding material
- f - frequency
- δ - depth of penetration

For simplicity, it is possible to determine the shielding effectiveness *SE* as:

$$SE = A + R$$

$$SE = 8{,}69.\frac{t}{\sqrt{\frac{2}{\omega.\mu.\sigma}}} + 20.\log\left(\frac{1}{4}.\sqrt{\frac{\sigma}{\omega.\mu_r.\varepsilon_0}}\right) \quad (2)$$

kde

- t - material thickness
- μ - permeability which is also included permeability shielding material
- σ - conductivity of shielding material
- μ_R - permeability of shielding material
- ε_0 - permittivity of vacuum

Equation (1) shows that the shielding effectiveness *SE* depend on the relative permeability and relative conductivity.

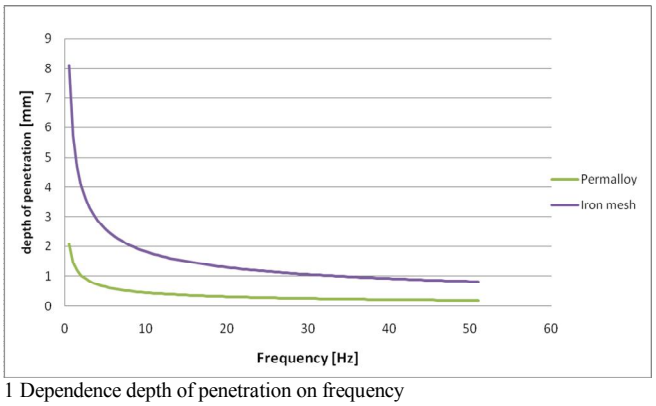


Fig. 1 Dependence depth of penetration on frequency

Fig. 1 shows the penetration depth dependence on frequency. The dependence is shown for frequencies up to 51 Hz. From dependence that with increasing frequency, the penetration depth decreases. This statement follows from regard to determine the depth of penetration:

$$\delta = \sqrt{\frac{2}{\omega.\mu.\sigma}} \tag{3}$$

where:

σ - conductivity of shielding material

μ - permeability which is also included permeability shielding material

f - frequency

δ - depth of penetration

This downward trend has also dependence on higher frequency.

Fig.2 and Fig.3 shows dependent absorption coefficient A and reflection coefficient R on frequency.

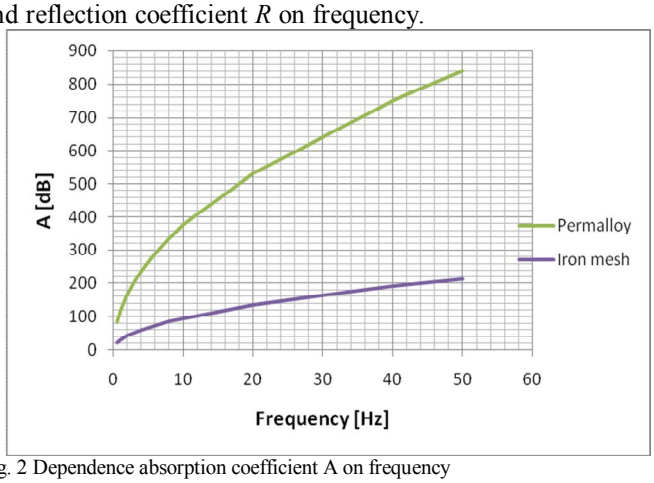


Fig. 2 Dependence absorption coefficient A on frequency

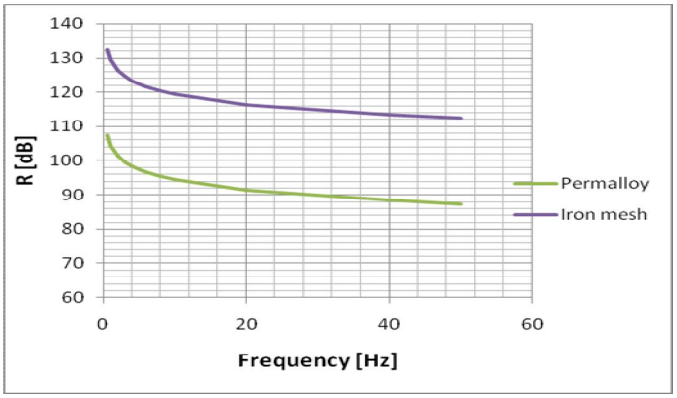


Fig. 3 Dependence reflection coefficient R on frequency

According to equation (2) can be derived dependence shielding effectiveness, depending on the frequency. This dependence is shown in Fig.4.

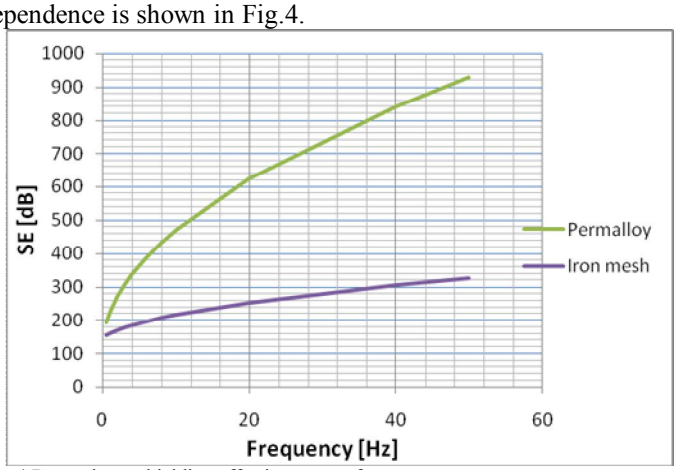


Fig. 4 Dependence shielding effectiveness on frequency

III. WORKPLACE FOR THE PURPOSE OF MEASURING SHIELDING EFFECTIVENESS OF ELECTROMAGNETIC FIELD

To determine the shielding effectiveness measurement is necessary to choose the material, which shielding effectiveness measurements we find. In this case, the material of the door shielding chamber at the Department of Electrical Power Engineering. Consequently, it is necessary to have a source of electromagnetic radiation. Sources of electromagnetic radiation as mobile phones, WiFi devices, microwave and so on.

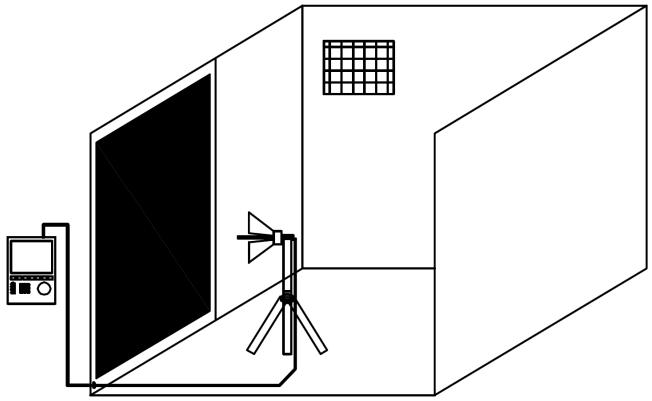


Fig. 5 Overall view of the measurement chamber shielding shield

The device, which shows the electromagnetic spectrum is a handheld spectrum analyzer Rodhe & Schwarz FSH8 (Fig. 6). The spectrum analyzer was connected to the receiving antenna Rodhe & Schwarz HF907 Horn (Fig 7), which was placed on a tripod. Overall view of the measurement chamber shielding shield is on Fig.5

Fig. 6 Spectrum analyzer R&S FSH8

Handheld FSH8 compact analyzer is designed to work in the laboratory for high-frequency devices. [4]

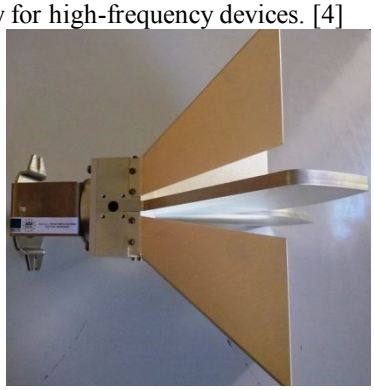


Fig. 7 Receiving Antenna R&S HF907 Horn

IV. RESULTS

In the introduction have been discussed frequency ranges of mobile network. This measure was focused on frequency ranges of 1, 2 and 3. Shielding effectiveness measured at frequency from 880MHz to 970 MHz and from 1700 to 1900 MHz. The frequency was measured in the range of 880 to 970 MHz in steps of 5 MHz to 900 MHz frequency monitored and between 1700 to 1900 MHz in increments of 5 MHz to 1800 MHz frequency. Measurements in the frequency range with this step, and for these two frequency range was made three times. The first measurement took place without shielding. shielding chamber door was open. Second measurement was made with shield, which consisted of iron mesh that was in the area of the door. A third measurement was made, when the door of shielding chamber was closed, resp. permalloy of which chamber is made.

Curves the electromagnetic field in dB are shown in Fig.8 for the frequency range of 900 MHz and Fig.9 for the frequency range of 1800 MHz. The graphs show that the iron mesh and permalloy shade these frequencies. The graphs also show that the frequency permalloy shade more than the iron mesh. This is due to the fact that while permalloy was formed fixed doors, iron mesh is made by mesh. If it were iron mesh as a whole, shielding would be higher.

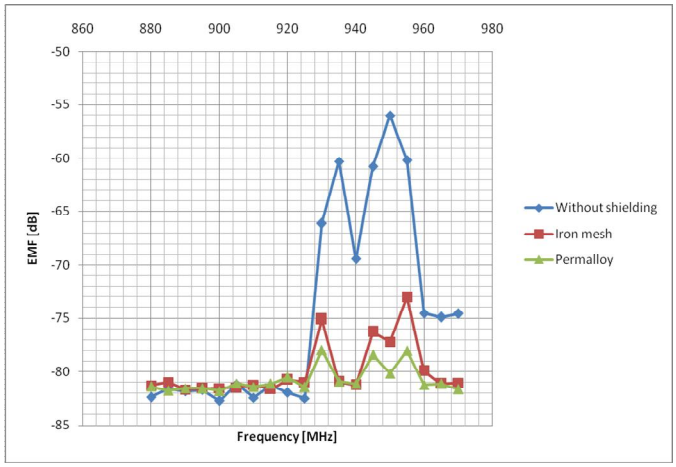


Fig. 8 Electromagnetic field for 900 MHz

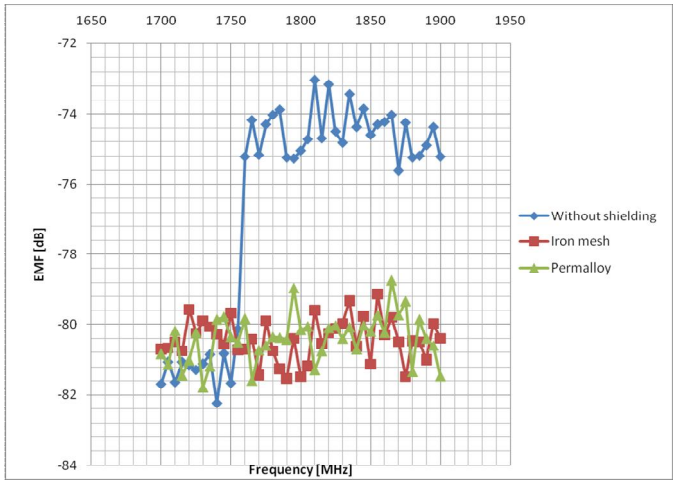


Fig. 9 Electromagnetic field for 1800 MHz

According to [5], it is possible to determine the shielding effectiveness *SE* curve for the permalloy (Fir.10) and iron mesh (Fig.11).

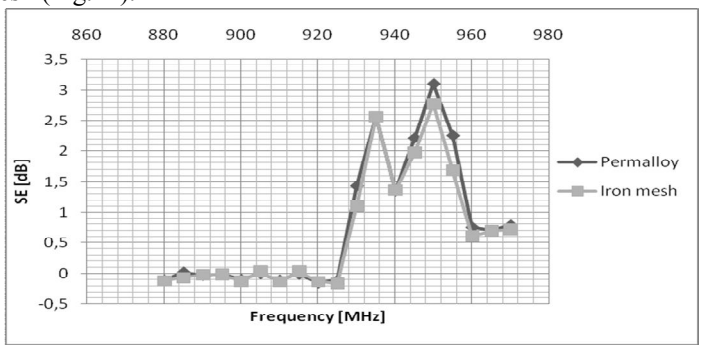


Fig. 10 Dependence of shielding effectiveness SE on frequency for 900 MHz

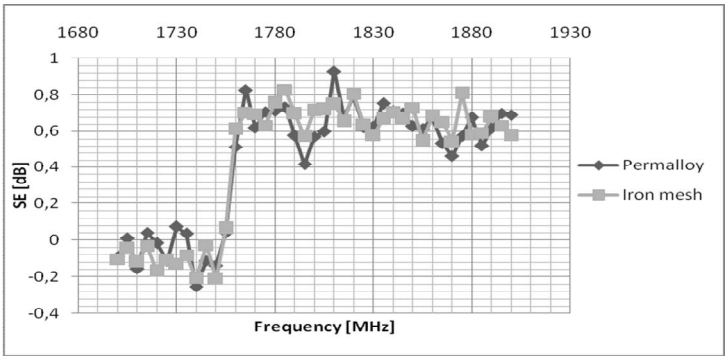


Fig. 11 Dependence of shielding effectiveness SE on frequency for 1800 MHz

V. CONCLUSION

This article deals with determination of shielding properties of shielding chamber. Shielding chamber consisted of material permalloy and iron mesh. The measurements show that as permalloy and iron mesh effectively shade electromagnetic field with a frequency range of 900MHz and 1800MHz. Measurement error may have occurred because as the measurement was made inside the chamber and occurred also reflections inside the chamber.

Shielding effectiveness *SE* of electromagnetic fields and electromagnetic compatibility is now becoming increasingly discussed topic, because of the continuous increase of electromagnetic field sources.

ACKNOWLEDGMENT

This publication is the result of the Project implementation: Protection of the population of the Slovak Republic from the effects of electromagnetic fields, ITMS: 26220220145, Activity 2.1 supported by the Research & Development Operational Programme funded by the ERDF .

We support research activities in Slovakia / Project is cofinanced from EU funds.

REFERENCES

[1] Plawiak-Mowna A., Krawczyk A., Wireless Body Sensor Network-Fundamental Concepts and Apalication, Przeglad Electrotechniczny, 12b/2012, 2012, s.267-268.

[2] Pan Z., Zhang H., et al.: Advances of Studies on Electromagnetism Shielding Fabric, Science & Technology Review, 27(2009), 24, 86-91.

[3] Kmec M., Hvizdoš M., Skúšky digitálnych ochrán prístrojom Omicron CMC, Electrical Engineering and Informatics 3 : proceeding of the Faculty of Electrical Engineering and Informatics of the Technical University of Košice. - Košice : FEI TU, 2012 S. 705-710. - ISBN 978-80-553-0890-6

[4] Manuál for handheld Spectrum analyzer R&S FSH4/ R&S FSH8: Product Brochure English available on: < http://www2.rohde-schwarz.com/file_19039/FSH4_FSH8_bro_en.pdf >

[5] IEEE Standart, Method for Measuring the Effectiveness of Electromagnetic Shielding Enclosures, EMC Society, New York 2006, p.39.

Method for ranging error mitigation of ToA-based systems applied in relative localization

Vladimír CIPOV (4th year)
Supervisor: Ľubomír DOBOŠ

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

vladimir.cipov@tuke.sk, lubomir.dobos@tuke.sk

***Abstract*— In this paper we introduce the concept of the range-based positioning algorithm for the relative localization in combination with Time of Arrival (ToA) node distance estimation ranging technique. The time of arrival of the received signal is determined using a peak detection algorithm from the received channel power delay profile. The node communication is simulated using the existing model for the channel impulse response (CIR) generation of the ultrawideband technology – UWB 802.15.4a. We describe the procedure for wireless channel identification which is very crucial for the precise ToA estimation. Additionally we propose the method of the ToA estimation error mitigation in the case of Non Line-of-Sight communication especially in the case of direct path unavailability.**

***Keywords*—UDP mitigation, ToA, anchor-free, positioning.**

I. INTRODUCTION

In the field of mobile communication, the wireless nodes location problem has gained increasing interest. Currently, the most challenging task for researchers is to solve the problem of anchor-free localization in Mobile Ad-hoc NETworks (MANETs) without any points creating the network infrastructure. From the available ranging techniques the ToA using UltraWideBand (UWB) communication technology is preferred, because it is one of the most promising approaches of the mobile positioning and has a great potential for accurate ranging especially in the harsh indoor environment. Due to very wide bandwidth of the UWB systems, these systems are capable resolving individual multipath components in the received CIR and determining their time of their arrival very precisely. We propose a low-complexity method for mitigation of the Undetected Direct Path (UDP) effect which causes considerable errors in the node distance estimation of the ToA-based systems. The method combines the mean of the time delay differences of the reflected paths detected in the measured channel power delay profile, overall channel impulse response length and Mean Excess Delay calculated relative to the first detected path. The positioning algorithm has been simulated in LoS, NLoS(OLoS) and NLoS(UDP) communication conditions.

II. THE TIME OF ARRIVAL ESTIMATION PROBLEM

Time of Arrival is the ranging technique where the node distance can be estimated simply as the product of the measured signal propagation delay ToA and the speed of light [1]. The conditions mostly influencing the node distance estimation accuracy of ToA ranging techniques are precision of transmitter-receiver time synchronization, used system band-width, multipath signal propagation and Non Line of Sight communication. The ranging error caused by multipath propagation is inversely proportional to the system bandwidth [2]. It is generally true that when the bandwidth of the system increases, the time-domain resolution and thus the accuracy of ToA estimation and node distance determination increase. In [3] the communication environment is classified as:

Line-of-Sight with Dominant-Direct-Path (LoS-DDP). In this case the direct path is the strongest and the first detected path in the channel profile and it can be detected by the receiver. There are no considerable obstacles between the transmitter and the receiver. However, the indoor radio propagation channel is characterized as exhibiting severe multipath propagation and low probability of LoS signal propagation between the transmitter and receiver [2].

Non Line-of-Sight with Non Dominant-Direct-Path (NLoS-NDDP), named also as Obstructed LoS (OLoS). Typical LoS communication between a transmitter and a receiver does not exist. In NDDP channel profile, the direct path is not the strongest path of the CIR, however, it is still detectable because it is received above the detection threshold of the receiver.

Non Line-of-Sight with Undetected-Direct-Path (NLoS-UDP). In UDP channels, the direct path goes below the detection threshold of the receiver while other paths are still detectable. The receiver assumes the first detected peak as a direct path which causes considerable distance estimation errors and consequently increasing of the node distance determination error. In the case when UDP is identified, the next step is to find ways, how to resolve the assumed direct path arrival time from the received channel power delay profile in order to minimize the ranging error.

At first the LoS/NLoS channel identification should be performed [5]. In order to identify the channel conditions the statistical Kurtosis Parameter (KP) can be used [4]. It is defined as the ratio of the fourth order moment of the data to the square of the second order moment. For the received CIR the value of KP is calculated. According to the set detection threshold, which represents the limit between the LoS and NLoS environment, the channel identification can be performed. If KP takes the values above the defined threshold it goes of LoS condition and vice versa, if it takes the values

less than the defined threshold it goes of NLoS condition. However, identification using only NLoS occurrence rather than UDP identification is not enough. Two UDP identification approaches that use the binary hypothesis testing of the defined channel parameters and an application of neural network architecture (NNA) design have been proposed [2].

III. Brief Overview of the Positioning Algorithm

The localization algorithm consists of two main stages:

I. **Formation of the Local Coordinate Systems (LCSs)**, where the Improved MDS-MAP technique is utilized in order to determine the relative node positions in LCS [6]. Two steps precede the MDS procedure.
   - **Node distance estimation** based on the ranging techniques such as Received Signal Strength (RSS) or Time of Arrival (ToA).
   - **Formation of the matrices containing the neighbouring node distances (DM)** derived in the previous step from the measured data.

II. **Formation of the Global Coordinate System (GCS)**, This process is also called the Network Coordinate System building [7] in where the partial Local Coordinate Systems are connected and unified in the same direction.

The aim of the process I. is formation of partial LCSs belonging to each node in the network. One LCS consists of the actual node relative position (this node is located at the centre of its LCS where its coordinates are set to [0,0]) and relative positions of its one-hop adjacent nodes. After the desired ranging technique procedure each node is able to fill up the first row and column of its *distance matrix (DM)*, where the cells represent the pair of the particular nodes, because it knows its distances to the one-hop adjacent nodes. Then the missing node distances are filled up in the DM. This process is based on the node communication procedure of such nodes, which belong into the LCS of the given node [8]. The DM contains the node pair distances, is symmetrical and has zeros in the main diagonal. The formed distance matrix is used as the input to the MDS method in order to form the LCS of each node.

Now, the network node LCSs are formed, but they have different directions. The partial LCSs must be joined together, because they are *shifted*, *rotated* and *mirrored* according to each other. Using such operations, in the second stage II. the partial LCSs can be joined into the one GCS. The procedure of the GCS formation is explained using two LCSs of the nodes A and B, Fig. 1. The LCS(B) is joined with the LCS(A). Presence of the node B located in the LCS(A) (and vice versa, the node A located in the LCS(B)) is crucial. Similarly, it is advantageous, if the third common node is located in LCS(A) and LCS(B) simultaneously which is needed in process of mirroring. At first the centre of the LCS(B) must be shifted to the node A, Fig. 2. The procedure of **shifting** is carried out by subtraction of the coordinates of the node A from the coordinates of each node. Next, both LCSs must be oriented in the same direction, in other words, LCS(A) must be **rotated** by an angle α and LCS(B) must be rotated by an angle β. They can be expressed for positive and negative y_B,y_B‘ respectively as, Fig. 3:

$$\alpha = \arccos\frac{x_B}{D_{AB}};\quad \beta = \arccos\frac{x_B^{'}}{D_{AB}^{'}} \qquad \alpha = -\arccos\frac{x_B}{D_{AB}};\quad \beta = -\arccos\frac{x_B^{'}}{D_{AB}^{'}} \qquad (1)$$

$[x_B,y_B]$ represent the coordinates of the node B and D_{AB} represents the distance between the nodes A and B. Similarly for $[x_B',y_B']$ and D_{AB}'. Then both LCSs will rotate in the direction negative to the size of α and β. For this process, it is necessary to calculate the angle δ, which represents the angle BAX_i (X_i represents *i-th* node) and can be calculated using the similar procedure to the previous angel calculation. The point coordinates will be recalculated for positive and negative result of δ- α as:

$$x_i = D_{Ai}\cos(\delta_i - \alpha);\quad y_i = D_{Ai}\sin(\delta_i - \alpha) \qquad x_i = -D_{Ai}\cos(\delta_i - \alpha);\quad y_i = D_{Ai}\sin(\delta_i - \alpha) \qquad (2)$$

The same procedure can be applied also for the rotation of the LCS(B). After rotation, a point C, which is common for both LCSs, must be found. According to its location, it will be decided about the need of **mirroring,** Fig. 4. If C and C’ have not the same sign of the y-axis, mirroring is necessary. It is carried out by simple change of the sign of the y-axis for each node in LCS. Now both LCSs can be joined together, Fig. 5.

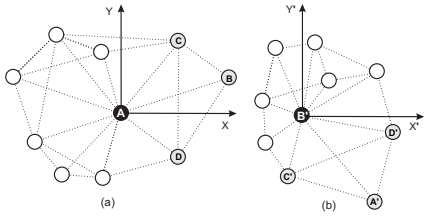


Fig. 1 LCSs of the nodes A and B used for problem explanation

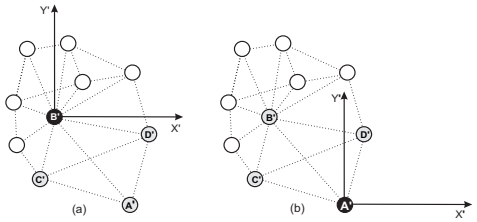


Fig. 2 LCS(B) shifted to the node A ((a)-before and (b)-after shifting respectively)

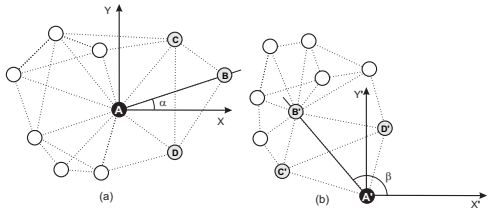


Fig. 3 Calculation of angles for rotation process

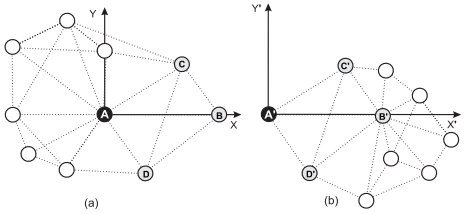


Fig. 4 LCSs oriented in the same direction

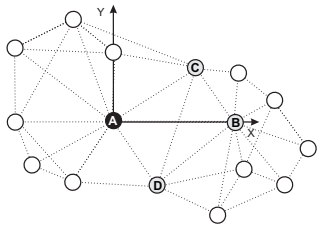


Fig. 5 LCS(B) joined to the LCS(A)

IV. Simulations and Results

In our simulations we have used the existing model of the wireless channel impulse response generation for IEEE 802.15.4a UWB technology [9]. We have utilized the generation of the CIRs for LoS and NLoS OFFICE environment. Our simulations have been based on a prior knowledge of the communication conditions without any channel identification procedure. In the most cases the ToA of the received signal is assumed as the ToA of the first detected path. The problem occurs, when the direct path is missing in the received signal CIR. This fact causes considerable distance estimation errors, witch leads to the incorrect localization. For NLoS mitigation problem, especially in case of NLoS communication with UDP, the low-complexity NLoS mitigation technique for ToA estimation has been proposed and it is described in the following part.

A. Description of the Proposed Low-Complexity method for UDP mitigation

The description of the method is explained using the concrete example of the transmitter-receiver pair 10m distant. Assume the CIR received at the receiver site as is depicted in Fig. 6 Detection threshold represents the sensitivity of the receiver and it is set to 50mV. The infinite dynamic range of the receiver has been assumed. All paths (multipath components) received above this detection threshold can be used for ToA estimation. It can be seen that in the received CIR the direct path is missing and it is necessary to estimate its assumed time of arrival $ToA_{assumed\text{-}LoS}$. At first the parameter, *Correction parameter* – C_{LoS}, must be calculated.

$$C_{LoS} = \left(\frac{TD_{mean}}{d(CIR)} \right) \cdot \tau_m \qquad (3)$$

It will be subsequently subtracted from the time of arrival of the first detected path $ToA_{Lp(1)}$ representing the first multipath component. TD_{mean} represents the mean of the time differences between the pairs of the paths received above the detection threshold. $L_{p(max)}$ denotes the serial number of the last received path and number of received paths simultaneously, $L_{p(i)}$ the serial number of the i-th received path. *d(CIR)* represents the total length of the CIR that is expressed as the difference between the ToA of the last received multipath component $ToA_{Lp(max)}$ and the ToA of the first multipath component $ToA_{Lp(1)}$. τ_m is the Mean Excess Delay calculated relative to the first detected path.

$$TD_{mean} = \frac{\sum_{i=2}^{L_{p(\max)}} (ToA_i - ToA_{i-1})}{L_{p(\max)} - 1} \qquad (4)$$

$$d(CIR) = ToA_{Lp(\max)} - ToA_{Lp(1)} \qquad (5)$$

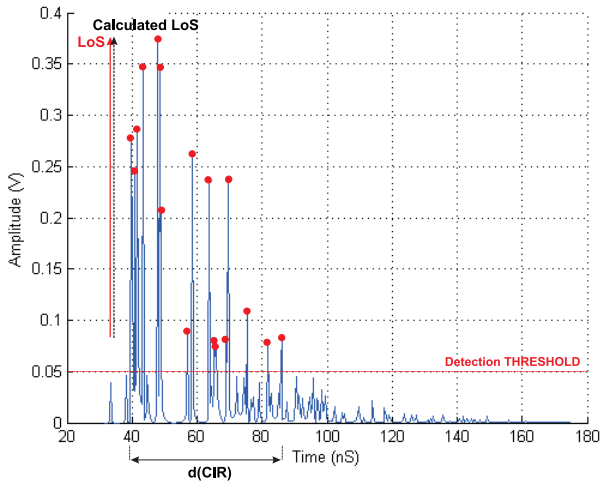


Fig. 6 CIR received at the receiver distant from the transmitter at 10m corrupted by UDP condition

In the next expression P_i represents the power of i-th received path and ToA_i its time of arrival.

$$\tau_m = \frac{\sum_{i=1}^{L_{P(\max)}} ToA_i |P_i|^2}{\sum_{i=1}^{L_{P(\max)}} |P_i|^2} \qquad (6)$$

After the calculation of the *Correction parameter* the assumed time of arrival of the direct path can be expressed as:

$$ToA_{assumed-LoS} = ToA_{Lp(1)} - C_{LoS} . \qquad (7)$$

In TABLE I the parameters of the received multipath components corresponding with the depicted CIR in Fig. 6 are introduced. The calculated parameters for *Correction parameter* calculation have the following values:

- $\boldsymbol{TD_{mean}}$ = 3,833ns; $\boldsymbol{d(CIR)}$ = 47,667ns; $\boldsymbol{\tau_m}$ = 50,191ns;
- $\boldsymbol{C_{LoS}}$ = 4,036 ns;

The ToA and node distance estimation has been improved as is expressed below:

- $\boldsymbol{ToA_{Lp(1)}}$ = 38,333ns; **Node distance** = 11,49m;
- $\boldsymbol{ToA_{assumed\text{-}LoS}}$ **= 34,297ns**; **Node distance = 10,289m**.

TABLE I
Power, Time of Arrival and Time Differences of the Received Multipath Components

i	P_i[V]	ToA_i[ns]	$ToA_i - ToA_{i-1}$[ns]
1	0.0823	38,3333	---
2	0.2774	39,6666	1,3333
3	0.2445	41,0000	1,3333
4	0.2858	41,6666	0,6667
5	0.3468	43,3333	1,6667
6	0.3739	48,0000	4,6667
7	0.3455	48,6666	0,6667
8	0.0893	57,0000	8,3334
9	0.2611	58,6666	10,0000
10	0.2359	63,6666	5,0000
11	0.0776	65,3333	1,6667
12	0.0740	66,0000	0,6667
13	0.0805	69,0000	3,0000
14	0.2355	69,6666	6,0000
15	0.1062	75,3333	5,6667
16	0.0749	81,6666	6,3333
17	0.0796	86,0000	4,3334

B. Obtained Results of the Proposed Method in Process of Distance Estimation and After its Implementation to the Positioning Algorithm

The proposed method has been tested on 1000-times randomly generated CIRs for transmitter-receiver pairs distant randomly at the distance from 0m to 15m. Only the harsh NLoS-UDP condition has been assumed. Generally, the distance estimation error has been reduced by 35,78%. The estimated ToA of the direct path has been improved in 74,5% of cases. In the TABLE II, the comparison of the distance estimation error using ToA of the first detected path and the assumed calculated ToA of the missing direct path can be found respectively.

TABLE II
DISTANCE ESTIMATION ERROR FOR NODES RANDOMLY DISTANT AT 0m - 15m

ToA estimation method	Distance estimation error [m]
$ToA_{Lp(1)}$	1,4356
$ToA_{assumed\text{-}LoS}$	0,9220

The proposed ToA estimation method has been tested also in Anchor-free positioning algorithm described in the section III. The impact of the proposed NLoS-UDP mitigation method on the potential enhancement of the localization error using this algorithm has been studied. For completeness, also the LoS and NLoS (Obstructed Line-of-Sight) node communication conditions have been simulated. The obtained results are introduced in TABLE III.

Our simulations were carried out in MATLAB programming environment and for statistical relevance were repeated one hundred times. The network consisting of the 50 mobile nodes deployed on the area of 50x50m with all receivers' radio range of 15m has been used.

The proposed UDP mitigation method has achieved the reduction of the relative localization error after the first stage of the algorithm, LCSs formation, of 18,11% and total relative distance estimation error of 35,61%. However we have expected the better result of the total localization error improvement. Despite the fact that the significant improvement of the node distance estimation has been reached, it is not sufficient for plausible reduction of the total localization error of the presented algorithm. Only small improvement of the localization accuracy has been enrolled. Note, if the localization error is expressed in %, we assume that error of 100% represent the value of the network size

TABLE III
OBTAINED RESULTS OF THE LOCALIZATION PROCESS AFTER IMPLEMENTATION OF THE PROPOSED UDP MITIGATION METHOD TO THE ANCHOR-FREE POSITIONING ALGORITHM

Type of office environment	Total Localization error (GCS) [%]	Localization error after 1st stage (LCS) [%]	Node distance estimation error [m]	Number of successfully localized nodes
NLoS-UDP	19,89	3,89	0,953	47,23
	20,41	4,75	1,480	47,64
NLoS (OLoS)	3,51	0,21	0,014	47,85
LoS-DDP	0,44	0,03	0,002	47,45

50m. The results achieved after the implementation of the proposed method of the UDP mitigation are marked grey.

V. CONCLUSION

The first simulations of the proposed low complexity method for node distance estimation enhancement using Time of Arrival are presented in this paper. The article is focused on the ToA estimation problem in harsh UDP environment where direct path is unavailable and on potential improvement of its ToA determination. The simulation results of the node distance estimation technique seem to be promising, but it is worth mentioning, that the simulations have been carried out in ideal communication conditions, especially with the precise time synchronization of mobile terminals which is very difficult to achieve in practice. Additionally, in the case of direct path presence (LoS/OLoS case), the model assumes the precise arrival time of the first detected path. These assumptions have affected the achieved simulation results that consequentially seem to be better than in reality. As a future work, it would be useful to compare the simulations with the measurements of the channel impulse response that suffer from the impact of the real communication environment

ACKNOWLEDGMENT

This work has been performed partially in the framework of the EU ICT Project INDECT (FP7 - 218086) (50%) and under research project ITMS-26220220155 supported by the Research & Development Operational Programme funded by the ERDF (50%).

REFERENCES

[1] N. Decarli, D. Dardari, S. Geizici, A.A. D'amico, "LOS/NLOS Detection for UWB Sig-nals: A Comparative Study Using Experimental Data," *IEEE International Symposium on Wireless Pervasive Computing (ISWPC)*, Modena, Italy, 5.-7. May 2010.

[2] M. Heidari, N. A. Alsindi, K. Pahlavan, "UDP identification and error mitigation in toa-based indoor localization systems using neural network architecture," Wireless Communi-cations, IEEE Transactions on , vol.8, no.7, July 2009, pp.3597-3607.

[3] N. Alsindi, "Performance of TOA Estimation Algorithms in Different Indoor Multipath Conditions," Diploma work. Faculty of Worcester Polytechnic Institute, Department of Electrical and Computer Engineering, Worcester, April 2004, 111p.

[4] I. Güvenc, Ch. Ch. Chong, F. Watanabe, "NLOS Identification and Mitigation for UWB Localization Systems," *Wireless Communications and Networking Conference, 2007. WCNC 2007*, IEEE. vol., no., 11.-15. March 2007, pp.1571-1576.

[5] N. Alsindi, C. Duan, T. Jinyun Zhang Tsuboi, "NLOS channel identification and mitigation in Ultra Wideband ToA-based Wireless Sensor Networks," *Positioning, Navigation and Communication, 2009. WPNC 2009*. 6th Workshop on , vol., no., 19. March 2009, pp.59-66.

[6] Y. Shang, W. Ruml, "Improved MDS-based localization," In: *Twenty-third Annual Joint Conference of the IEEE Computer and Communications Societies, INFOCOM 2004*, vol.4, Hong Kong, 7.-11. March 2004, pp. 2640 – 2651.

[7] S. Čapkun, M. Hamdi, J.P. Hubaux, "GPS-free positioning in mobile ad hoc networks," In: Cluster Computing, vol. 5 (2), April, 2002, pp. 157 – 167.

[8] V. Cipov, M. Copák, "Anchor-free positioning using ToA node distance estimation and LoSNLoS detection," In: *SCYR 2012*: Proceedings from conference: *12th Scientific Conference of Young Researchers*, Herľany, Slovakia. - Košice : TU, 15. May 2012, pp. 23-26, ISBN 978-80-553-0943-9.

[9] A.F. Molisch, et al., "IEEE 802.15.4a Channel Model – Final Report" Tech. Rep., Document IEEE 802.1504-0062-02-004a, 2005.

Modular object tracking system using CPU, GPU or FPGA units

[1]*Ján VALISKA (2^{st} year),* [2]*Ondrej KOVÁČ (2^{nd} year)*
Supervisor: [3]*Stanislav MARCHEVSKÝ*

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

[1]jan.valiska@tuke.sk, [2]ondrej.kovac@tuke.sk, [3]stanislav.marchevsky@tuke.sk

***Abstract*—In this paper, modular system for object tracking in video sequences using Bayesian filters based on tracking of color distribution of objects is proposed. Modularity of system allows to define different inputs and describe their dependencies and properties. Since calculating or estimating object position with selected algorithm of particle filter is complex task, system allows to utilize computation threads in multiple units. At first, there is CPU processor, that at higher spatial resolution of video sequence does not provide sufficient power. At second, system allows usage of GPU or FPGA units, which provide sufficient power for many tracking tasks. Modularity of proposed system allows to develop new tracking kernel and is suitable as a basis for development and research.**

***Keywords*—FPGA, GPU, histogram, object tracking, video**

I. Introduction

Object tracking is a task needed by many video applications, like human to PC interface, video communications, video compression, road-traffic control or video surveillance systems. Frequently requested output from tracking application is trajectory of moving objects, that depends on time and space, obtained by several sensors e.g. video camera.

In this computing task, processing a large amount of data from sensors is needed, which is very difficult to compute and has big demands on computing system. This paper deals with theoretical dissection of multiple object tracking using particle filters and proposal of complete modular system for real-time multiple object tracking. Proposed system is also able to classify and identify objects using patterns in database and insert data about tracked objects into a database.

This system should be capable of replacing human operators meanwhile video-surveillance task, by replacing partial function of the operator. Specifically, system tries to detect, identify, and classify the object and track it by object priority.

In the first section particle filter for video-object tracking using color distribution of object (histogram) is proposed. In the second section, system for handling multiple inputs (cameras) using CPU, GPU and FPGA units is proposed. Afterwards, system kernel and process of interaction between human and this system are described.

II. Proposal of filter using color distribution

Particle filtering was developed for object tracking tasks in dynamic and quickly alternating video scenes, where posterior density $p(X_t|Z_t)$ and observe density $p(Z_t|X_t)$ are often non-Gaussian. Tracking objects quantity is described in state vector X_t and vector Z_t consist of all measurements $\{z_1, \ldots, z_t\}$ up to time t.

Key feature of particle filtering is approximation of probability density using weighted sample set $S = \{(s^{(n)}, \phi^{(n)})|n = 1, \ldots, N\}$. Every particle consist of element s, describing hypothetical state of object and element a describing discrete sampling probability π where $\sum_{n=1}^{N} \pi^{(n)} = 1$.

Sample set evolution is produced by processing particles using system model. Next are every element of sample set weighted using measurement rule and N particles are drawn and replaced by particular samples with probability $\pi^{(n)} = p(z_t|X_t = s_t^{(n)})$. Mean state of object is estimated in every time by

$$E[S] = \sum_{n=1}^{N} \pi^{(n)} s^{(n)}. \quad (1)$$

Particle filtering provides a robust tracking base, due to the instability of the object model. The filter allows to have open space for object properties and evaluate their various hypotheses. Since state of object is relatively continuous in time while tracking process, particle filters are able to deal with short outage of tracking. [1]

A. Color distribution model

The proposed filter is adapted to track color distribution of object. To achieve robustness against image flexibility, rotations and particle occlusions, is for the filter selected state model based on color distribution of the image or object. These states are represented by histograms, which are obtained using $h(x_i)$, which assigns one m-bins for that color at position x_i. Histograms are typically calculated in the RGB space with 8x8x8 bins. For a more robust algorithm against lighting conditions may be used HSV model with less sensitivity to V(luminance) component, for example, using 8x8x4 bins.

Not all pixels in the viewing area are equally important to describe the object. For example, pixels that are placed in further from the center of the region, may have allocated smaller weights by applying weighting function:

$$k(r) = \begin{cases} 1 - r^2 & : \quad r < 1 \\ 0 & : \quad inak \end{cases} \quad (2)$$

where r is distance from center of region. That leads to more reliability in color distribution if these pixels belongs to background image. It is also possible use other weighting function e.g. Epanechnik kernel [2].

Color distribution $p(y) = \{p(y)^{(u)}_{u=1,\dots,m}\}$ at position y is calculated by:

$$p(y)^{(u)} = f \sum_{i=1}^{I} k\left(\frac{||y - x_i||}{a}\right) \delta[h(x_i) - u] \tag{3}$$

where δ is Kronecker delta function and parameter a is used for adjust region dimensions. Normalization factor

$$f = \frac{1}{\sum_{i=1}^{I} k\left(\frac{||y-x_i||}{a}\right)} \tag{4}$$

ensure that $\sum_{u=1}^{m} p(y)^{(u)} = 1$.

In tracking process is estimated state in every time updated by applying new measurements. Therefore is needed to use similarity measure, which is based on color distribution. Frequently used measure between two distributions $p(u)$ and $q(u)$ is "Bhattacharyya" coefficient [3], [4]

$$\rho[p, q] = \int \sqrt{p(u)q(u)}du. \tag{5}$$

Considering discrete densities such as our color histograms $p = \{p^{(u)}\}_{u=1,\dots,m}$ and $q = \{q^{(u)}\}_{u=1,\dots,m}$ the coefficient is defined as

$$\rho[p, q] = \sum_{u=1}^{m} \sqrt{p^{(u)}q^{(u)}}. \tag{6}$$

The larger ρ is, the more similar the distributions are. For two identical histograms we obtain $\rho = 1$, indicating a perfect match. As distance between two distributions we define the measure

$$d = \sqrt{1 - \rho[p, q]}, \tag{7}$$

which is called the Bhattacharyya distance.

B. Object tracking algorithm

The proposed tracker employs the Bhattacharyya distance to update the apriori distribution calculated by the particle filter. The target regions are represented by ellipses, so that a sample is given as

$$s = \{x, y, \dot{x}, \dot{y}, H_x, H_y, \dot{H}_x, \dot{H}_y\} \tag{8}$$

where x, y represents position of ellipse, $\hat{x}, \hat{y}$ representing motion directions, H_x and H_y are representing half-axis length, $\hat{H}_x$ and $\hat{H}_y$ corresponding changes of half-axis. As we consider a whole sample set the tracker handles multiple hypotheses simultaneously.

The sample set is propagated through the application of a dynamic model

$$s_t = As_{t-1} + w_{t-1} \tag{9}$$

where A defining deterministic and w_{t-1} stochastic element. In this application of filter is used first-order model A, that describes moving object with constant velocity for x, y, H_xandH_y. Expanding of model to second-order is straightforward.

To weigh the sample set, the Bhattacharyya coefficient has to be computed between the target distribution and the distribution of the hypotheses. Each hypothetical region is specified by its state vector $s^{(n)}$. Both the target q and candidate histogram $p(x^{(n)})$ are calculated from (3), where the target is centered in the origin $a = \sqrt{H_x^2 + H_y^2}$. In proposed system is used fixed target model, but it is also possible to implement any adaptive methods presented in [5], [6].

Fig. 1. One iteration step of filter

Get sample set S_{t-1} and target model $q = f\sum_{i=1}^{I} k\left(\frac{||x_i||}{a}\right)\delta[h(x_i) - u]$ and execute next steps:

1) **Select** N samples from set S_{t-1} with probability $\pi_{t-1}^{(n)}$:
   a) calculate the normalized cumulative probabilities $c_{t-1}^{'}$

$$\begin{aligned} c_{t-1}^{(0)} &= 0 \\ c_{t-1}^{(n)} &= c_{t-1}^{(n-1)} + \pi_{t-1}^{(n)} \\ c_{t-1}^{'(n)} &= \frac{c_{t-1}^{(n)}}{c_{t-1}^{(N)}} \end{aligned}$$

   b) generate a uniformly distributed random number $r \in [0, 1]$
   c) using binary search, find smallest j, for which $c_{t-1}^{'(j)} \geq r$
   d) set $s_{t-1}^{'(n)} = s_{t-1}^{(j)}$
2) **Propagete** each sample from the set $S_{t-1}^{'}$ by a linear stochastic differential equation:

$$s_t^{(n)} = As_{t-1}^{'(n)} + w_{t-1}^{(n)}$$

   , where $w_{t-1}^{(n)}$ is the stochastic component
3) **Observe** the color distribution:
   a) calculate the color distribution

$$p(s^{(n)})^{(u)} = f \sum_{i=1}^{I} k\left(\frac{||s_t^{(n)} - x_i||}{a}\right) \delta[h(x_i) - u]$$

   for each sample in set
   b) calculate the Bhattacharyya coefficient for each sample of the set S_t

$$\rho[p(s_t^{(n)}), q] = \sum_{u=1}^{m} \sqrt{p(s_t^{(n)})^{(u)} q^{(u)}}$$

   c) weight each sample

$$\pi_t^{(n)} = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{(1-\rho[p(s_t^{(n)}),q])}{2\sigma^2}}$$

4) **Estimate** the mean state of the set S_t

$$E[S_t] = \sum_{n=1}^{N} \pi_t^{(n)} s_t^{(n)}$$

As we want to favor samples whose color distributions are similar to the target model, the Bhattacharyya distance is used for the weighting. The probability of each sample

$$\pi^{(n)} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{d^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(1-\rho[p(s^{(n)})]),q}{2\sigma^2}} \tag{10}$$

is specified by a Gaussian with variance σ. During filtering, samples with a high weight may be chosen several times, leading to identical copies, while others with relatively low weights may not be chosen at all. The programming details for one iteration step are given in (1).

C. System initialization

Initialization of system consist of findings starting values x, y, H_x and H_y. There are three choices dependent on prior knowledge of target object: manual initialization, automatic initialization using known histogram as target model and by using algorithm for detection objects in video-sequences. Manual initialization require human operator to select object before tracking. Object must to be fully visible to calculate good color distribution.

If is known histogram of object $q = \{q_{u=1,\dots,m}^{(u)}\}$, we can

strategically place samples at positions with most probability of object occurrence.

The tracker should detect the object when it enters the field of view of the camera. In this case, the Bhattacharyya coefficient in the vicinity of the object position should be significantly higher than the average coefficient of the background. Therefore, we first calculate the mean value μ and the standard deviation σ of the Bhattacharyya coefficient for every pixel of a background image:

$$\mu = \frac{1}{M}\sum_{i=1}^{M}\rho[p(x_i), q] \tag{11}$$

$$\sigma^2 = \frac{1}{M}\sum_{i=1}^{M}(\rho[p(x_i), q] - \mu)^2 \tag{12}$$

and then define an appearance rule as

$$\rho[p(s_t^{(n)}), q] > \mu + 2\sigma \tag{13}$$

This indicates a 95% confidence that a sample does not belong to the background. If more than a fraction f of the sample set fulfills the appearance rule during initialization, we consider the object to be found and start tracking.

Likewise, the same rule is used to determine if an object is lost during the tracking. If the number of positive appearances is smaller than f for a couple of frames, the tracker enters the 'initialization' mode. [7]

III. Modular system design

In this section is proposed complete modular system for object tracking in video-sequences. Base block scheme is at figure (2).

Complete system for tracking must be able to select signal from multiple input sources(like IP, USB camera, or file). Because there are many of protocols used in video streaming, we select only HTTP protocol through IP network. System is proposed to be able to get signal also from USB camera, which is in present time used widespread. Selecting and processing function of input streams is processed by input interface, which is configurable by configuration file in XML standard.

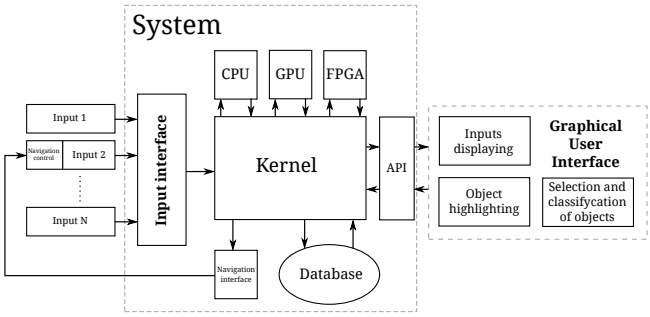


Fig. 2. System block scheme

Since cameras used in security systems often consist of servo-motor used to rotate objective of camera to target, proposed input interface must be able also to do this task, and also must describe all protocols of inputs witch using servo-motors.

For the purpose of testing, we proposed also simple protocol between PC and servo-motors of USB camera, which will be communicating using RS232 link. Alone control of camera will be implemented on development board based on Arduino project.

One of the challenges of object tracking is to make real-time system, which is able to analyze, detect and track objects. Biggest computational power consumption problem or task in object tracking by color distribution of objects, is processing huge amount of particles. Every object must have assigned sufficient number of particles, what may be more than 1000. This problem is solvable by using close set of object properties, using fast algorithms for measurements and similarity measures of particles, or using fast computational devices like GPU or FPGA units. Since proposed system must be modular, it should be have dynamic reactions to change in video frames, number of objects, objects priorities and automatically assign number of computational threads to every object. According to object priorities and difficulty, system should be select best and effective computational method(CPU, GPU or FPGA).

That functionality will ensure a system kernel, which after successful detection of object evaluate object and assign computational threads for CPU, GPU or FPGA units. On every unit, will be created static numbers of threads, which run will be started after request of system kernel. Kernel ensures also gathering informations data from objects to computational threads and also processing threads outputs. Kernel consist also of object register with all informations about them. These include the actual positions, attributes, and properties of objects. Interface for camera navigation gather informations about objects from system kernel, and after evaluating this information, this interface navigate cameras(with servo-motor) in direction of moving object. System kernel must be able in every time to compensate changes in background image of camera.

One of the periphery of the core is the API through which will provide access to the core functions. The API that will be connected via a graphical interface written in QT. The interface will be possible to select and view the various inputs available (images from the cameras). If the kernel detects a new object (person), so it enhances the graphical user interface and consequently, if the object is recognized as a known (the database), lists the name of the object or its properties. If the object is unknown, will have a 'User' option have selected for the object, save it to the database and assign it a name and a variety of features (symptoms). The system should be capable of providing access to the core by other user interface (textual interface or HTTP web-based interface for remote access).

The detection of persons is ensured by Haar-cascade classifier algorithm, configured for the detection of human body. Subsequently, the algorithm tries in the upper section of object region detect face (also by cascade classifier). After successfully detecting the faces extracted from the color information (histogram), will be that object selected as target object. After this step, the system classifies this object and start tracking in a new tracking thread and allocate free computing threads in CPU, GPU or FPGA units. Throughout the monitoring process are measurement and consequently estimated values()positions) stored in the database in order to access this information in the future (e.g. forensic analysis). The database will also store a whole histograms of detected objects and objects screen shots from different angles, to allow the detection and classification of persons for future occurrences.

Because of access speed to database are database data on

initialization of system transfered to RAM memory, what leads to growth of system resources, because of read and write speed are approximately ten times faster in system RAM memory. To compare reading and writing speed using MySQL database, located in RAM file system, I present these results:

Read and write on HDD

```
OLTP test statistics:
    queries performed:
        read:                            1400056
        write:                           500020
        other:                           200008
        total:                           2100084
    transactions:                        100004 (111.50 per sec.)
    deadlocks:                           0      (0.00 per sec.)
    read/write requests:                 1900076 (2118.44 per sec.)
    other operations:                    200008 (222.99 per sec.)

Test execution summary:
    total time:                          896.9235s
    total number of events:              100004
    total time taken by event execution: 14349.6163
    per-request statistics:
         min:                                 49.65ms
         avg:                                143.49ms
         max:                               1532.38ms
         approx.  95 percentile:             275.55ms

Threads fairness:
    events (avg/stddev):           6250.2500/12.02
    execution time (avg/stddev):   896.8510/0.04
```

Read and write on RAM

```
OLTP test statistics:
    queries performed:
        read:                            1400000
        write:                           500000
        other:                           200000
        total:                           2100000
    transactions:                        100000 (1019.67 per sec.)
    deadlocks:                           0      (0.00 per sec.)
    read/write requests:                 1900000 (19373.71 per sec.)
    other operations:                    200000 (2039.34 per sec.)

Test execution summary:
    total time:                          98.0710s
    total number of events:              100000
    total time taken by event execution: 1568.5880
    per-request statistics:
         min:                                  2.44ms
         avg:                                 15.69ms
         max:                                536.66ms
         approx.  95 percentile:              25.52ms

Threads fairness:
    events (avg/stddev):           6250.0000/140.94
    execution time (avg/stddev):   98.0367/0.00
```

IV. Conclusion

The main advantage of proposed modular system is design itself. It allows to develop or research new detecting, tracking, or classifying algorithms. Interface for connecting the powerfull units like GPU or FPGA allows to bridge problem of high demands on computing power. Whole system can be used also in production for surveillance of buildings or open-air regions and also for work facilitation of human operators in security centers.

Acknowledgment

This work is the result of the project implementation: Development of the Center of Information and Communication Technologies for Knowledge Systems (ITMS project code: 26220120030) supported by the Research & Development Operational Program funded by the ERDF.

References

[1] K. Nummiaro, E. Koller-meier, and L. V. Gool, “A color-based particle filter,” 2002, pp. 53–60.

[2] D. Comaniciu, V. Ramesh, and P. Meer, “Real-time tracking of non-rigid objects using mean shift,” in *Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No.PR00662)*, vol. 2. IEEE Comput. Soc, 2000, pp. 142–149. [Online]. Available: http://dx.doi.org/10.1109/CVPR.2000.854761

[3] N. A. Thacker, F. J. Aherne, and P. I. Rockett, “The bhattacharyya metric as an absolute similarity measure for frequency coded data,” *Kybernetika*, vol. 34, no. 4, pp. 363–368, 1997. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.114.1002&rep=rep1&type=pdf

[4] T. Kailath, “The Divergence and Bhattacharyya Distance Measures in Signal Selection,” *Communication Technology, IEEE Transactions on*, vol. 15, no. 1, pp. 52–60, Feb. 1967. [Online]. Available: http://dx.doi.org/10.1109/tcom.1967.1089532

[5] A. D. Jepson, D. J. Fleet, T. F. El-maraghi, I. C. Society, I. C. Society, and I. C. Society, “Robust online appearance models for visual tracking,” 2001, pp. 415–422.

[6] Y. Raja, S. J. Mckenna, and S. Gong, “Tracking and segmenting people in varying lighting conditions using colour,” in *In AFG*, 1998.

[7] D. Comaniciu and V. Ramesh, “Mean shift and optimal prediction for efficient object tracking,” in *Tracking, International Conference on Image Processing*, 2000, pp. 70–73.

Motor acceleration analysis and synchronous generator speed governing behavior

[1]*Peter HERETIK (2st year),* [2]*Marian HALAJ (5st year)*
Supervisor: [3]*Justín MURIN, Ladislav VARGA*

[1,3]Institute of Power and Applied Electrical Engineering FEI STU, Slovak University of Technology, Ilkovičova 3, 812 19 Bratislava, Slovak Republic

[2,4]Dept. of Electrical Power Engineering, FEI TU of Košice, Slovak Republic

[1]peterheretik@gmail.com, [2]marian.halaj@centrum.sk, [3]justin.murin@stuba.sk, [4]ladislav.varga@tuke.sk

***Abstract*—The main purpose of this contribution is to examine properties of chosen electrical machines. Calculations and simulations are aimed to an asynchronous motor and a synchronous generator. For calculations are used data of these machines provided by real manufacturers as an input for program ETAP. Motor acceleration analysis is demonstrated according to given motor load characteristic in program ETAP. Simulation from ETAP is compared to manual calculation and to the value provided by manufacturer. Important issue for analyzing the synchronous generator is the device regulating the speed of the generator, the governor. Response of the generator to connecting loads is influenced by the reaction of the governor, which is examined in isochronous and droop mode.**

I. ETAP

ETAP (Electrical Transient Analyzer Program) is the most comprehensive analysis platform for the design, simulation, operation, and automation of generation, distribution, and industrial power systems. ETAP is developed under an established quality assurance program and is used worldwide as a high impact software. It is completely localized in four languages with translated output reports in six languages.

As a fully integrated enterprise solution, ETAP extends to a Real-Time Intelligent Power Management System to monitor, control, automate, simulate, and optimize the operation of power systems. [1]

ETAP offers a suite of fully integrated software solutions including arc flash, load flow, short-circuit, relay coordination, cable ampacity, transient stability, optimal power flow, voltage drop and more. [2]

II. PARAMETERS OF ELECTRICAL MACHINES

Parameters of motor and generator used for dynamic calculations in following chapters of this contribution are real data from manufacturers of specific electrical machine. These parameters are provided by manufacturer in datasheet, what is a document summarizing these parameters.

A. Induction motor

A typical datasheet for an induction motor contains information about rated values (voltage, power, speed, frequency, torque etc.), critical values which can not be exceeded or underrated (breakdown torque, minimal starting voltage). An example of parameters of induction motor necessary for dynamic calculations is in Table 1.

TABLE I
DATA OF AN INDUCTION MOTOR [4]

Description	Unit	Value
Component identification Code	-	RCV5130POM
Manufacturer	-	SCHORCH
Serial number of motor	-	44047811
Rated power	kW	480
Rated voltage Ur/ connection	kV	10 / star
Minimal starting voltage	% U_r	80
Rated speed	rpm	2975
Rated frequency	Hz	50
Rated efficiency	%	95
Efficiency at: 3/4 load 1/2 load	 % %	 95 94,5
Rated power factor cosφ	-	0,9

A torque/speed characteristic, supplied by manufacturer, is also necessary for dynamic calculations.

Values from Table I are used in program ETAP for Transient stability and Motor acceleration analysis.

B. Synchronous generator

Synchronous generator datasheet contains designed data like climate, important information about rated values as a voltage, power factor, efficiency etc. Data about type of generator that should be known for working with it, as a number of poles, direct-axis synchronous reactance, negative sequence resistance.

An example of data of a synchronous generator is in Table°II.

TABLE II
DATA OF A SYNCHRONOUS GENERATOR [5]

Description	Unit	Value
Manufacturer	-	JEUMONT Electric
Rated power	kW	480
Rated apparent power	kVA	7800
Rated voltage	kV	10 500
Rated current	A	429
Rated frequency	Hz	50
Rated power factor	-	0,8
Rated speed	rpm	1000
Rated efficiency	%	97,15
Number of poles	-	6

For dynamic calculations we also need generator capability curve. It shows relation between active and reactive power.

III. Dynamic Calculations

For motor acceleration analysis was used scheme in Figure I. In this scheme an asynchronous motor is supplied by power from a synchronous generator.

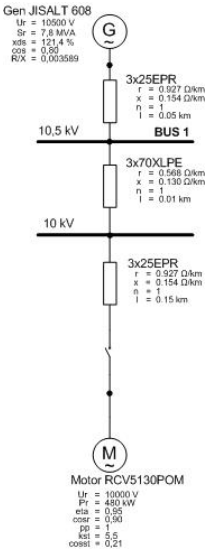


FIGURE I
SIMULATION SCHEME

A. Motor acceleration calculation

One of the most important values at starting an induction motor is the starting time at rated voltage U_r and rated load. This value should be provided by manufacturer in datasheet. Starting time provided by manufacturer of the motor used in Figure I is 2 seconds. This time can be simulated from time dependence of speed in program ETAP.

Time dependence of speed, terminal current and terminal voltage simulated by ETAP is in Figure II.

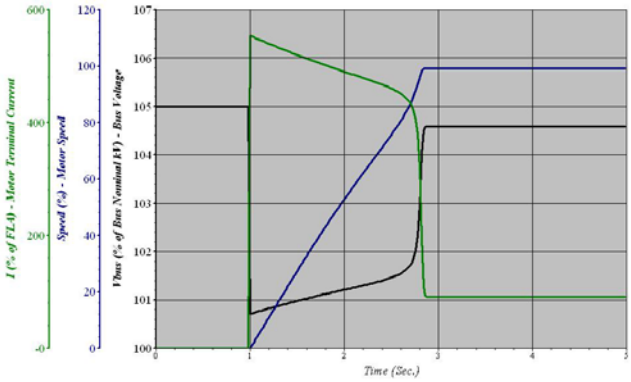


FIGURE II
TIME DEPENDENCE OF CURRENT, SPEED AND VOLTAGE

In Figure II can be seen that the motor was started in 1 second and reached its rated speed in about 2.8 seconds. The result of subtracting these two values, 1.8 seconds, is the starting time of the motor calculated by program ETAP.

Starting time can be calculated manually from motor torque/speed characteristic with motor torque and load torque curves. For calculation is needed acceleration torque curve which is the result of subtracting motor torque and load torque curves. The starting time can be calculated from following equation:

$$\Delta t = \frac{J}{M_a}.\Delta\omega \text{ [3]} \qquad (1)$$

where J is moment of inertia of motor and driven machine from datasheet, M_a is motor acceleration torque and $\Delta\omega$ is increment of the angular velocity of the motor. Moment of inertia J is provided by manufacturer in the datasheet. For motor used in Figure I, the value for moment of inertia is $J=6kg.m^2$.

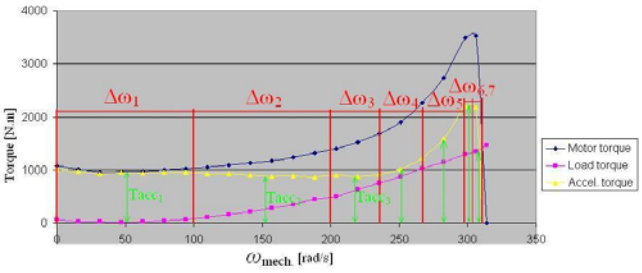


FIGURE III
ACCELERATION TIME CALCULATION

In Figure III can be seen the angular speed dependence of torque of the induction motor used in Figure I, which was created manually according to curves provided by manufacturer in the datasheet. The curve of acceleration torque is divided into n sections which are interpolated by straight lines and the mean value of torque for each interval $\Delta\omega_n$ is used to calculate the acceleration time in specific interval. Following equation represents the calculation for the first interval:

$$t_1 = \frac{6}{975}.100 = 0{,}61s \qquad (2)$$

The result of sum of all partial times is the acceleration time of the induction motor.

$$t = \sum_{1}^{n} \frac{J}{M_{a_n}} \Delta\omega_n = 1.857s \quad (3)$$

The starting time calculated manually is 1.857 seconds. It is clear, that both of starting times are smaller than the time provided by manufacturer, which means that the motor will accelerate in less than 2 seconds at rated voltage and rated load. Considering that the motor is powered by a generator with rated voltage 10.5 kV used in electrical scheme in FigureºI, the starting time calculated by ETAP could be smaller than the time from datasheet.

B. The Governor

This part describes the representation of speed governing and prime mover control systems for synchronous generators. The governor is a device used to measure and regulate the speed of a machine.

In isochronous speed control mode, energy being admitted to the prime mover is regulated tightly in response to changes in load which would tend to cause changes in speed, so the speed is still maintained by prime mover at the setpoint. For example see Figure IV. After connecting the single loads is still maintained the rated speed until the generator is overloaded with the last load.

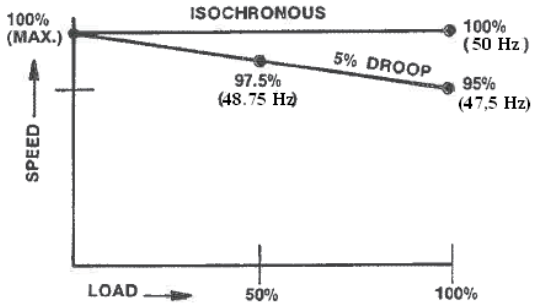


FIGURE IV
ISOCHRONOUS AND DROOP MODE DEFINITION

Droop mode is defined with a decrease in speed setting of generator as the load increases. Speed droop is a decrease in speed or frequency, proportional to load. If the load increases, then the speed or frequency decreases (drops). Speed droop is expressed as the percentage reduction in speed versus the speed setpoint that occurs when the generator is unloaded.

$$Droop = \frac{NoLoadSpeed - FullLoadSpeed}{NoLoadSpeed} \times 100 \quad (4)$$

In Figure IV the line is defined with speed at 0% load and speed at 100% load. The line shows ideal values of decreased speed at load from 0% to 100% load at defined droop. At 100% load the generator speed decreased from 1000 RPM to 925 RPM (see Figure VI). Difference between them is 7.5% that equals 3.75Hz of 50Hz rated frequency of the generator.

Relation between generator frequency and speed of generator:

$$n = \frac{f}{p} \quad (5)$$

Where n is speed in rpm, f is frequency in Hz and p is number of pairs of poles.

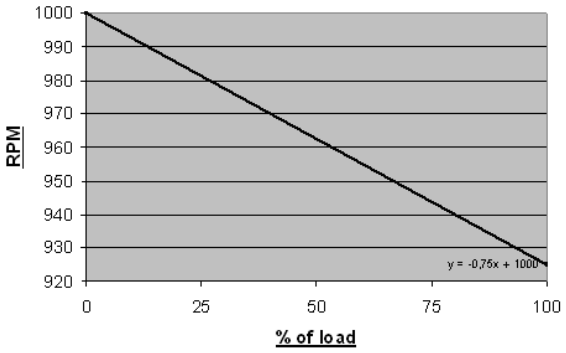


FIGURE V
SPEED DECREASE CHARACTERISTIC

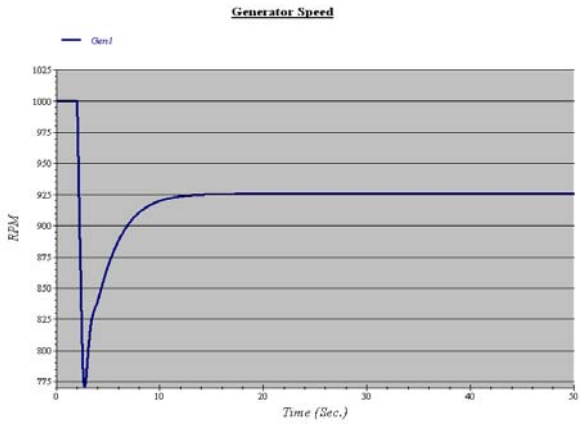


FIGURE VI
GENERATOR SPEED DECREASE AT RATED LOAD

In Table III are compared values of generator speed after application of load at 100%, 75%, 50% and 25% from the simulation and ideal values from graph Figure V. There can be seen negligible differences between compared values.

TABLE III
SPEED DROOP

Load	Simulation RPM	Speed decrease characteristic RPM
100%	925	925
75%	944,1	943,75
50%	962,5	962,5
25%	981,1	981,25

IV. Conclusion

In conclusion of calculating acceleration time of asynchronous motor can be shown, that manual calculations are correct according to calculations from ETAP and a deviation between them, which was 0.057 seconds, is negligible. In further studies of induction machines will be determined voltage ratios of individual motors and a group of motors at the starting, impact of the reactance of the transformer and generator on voltage drop at starting of asynchronous motors, eventually the short circuit analysis of induction motors.

In governor calculating is verified, that this device regulates speed of prime mover according to adjusted speed decrease characteristic. There are no noticeable speed decreases in the governor isochronous mode, on the other hand, operating with governor in droop mode shows speed decreases according to size of the connected load. In further

studies of synchronous generators will be closely analyzed excitation systems of generators, their prime movers responding to aspects of electrical grid, globally dynamic calculations of synchronous generators.

ACKNOWLEDGEMENTS

This paper has been supported by project CRISIS ITMS 26240220060.

I would like to thank Research described in the paper was supervised by Ing.ºMartin Naňo and Ing. Juraj Breza, PhD. and supported by AREVA NP Controls, s.r.o.

I would also to thank the supervisor of my PhD study prof.ºJustín Murín from IPAE FEI STU in Bratislava for the help by the preparation of the contribution.

REFERENCES

[1] ETAP Operation Technology, INC., ETAP 11.0.0, *Product overview*, B8-PO-E11-0911-15.
[2] *ETAP User Guide, version 1110. Germany, 2012*.
[3] HRUSKOVIČ, L. *Elektrické stroje*. Bratislava: Vydavatelstvo STU, 1999, 497p. ISBN 80-227-1249-3.
[4] Electrical data provided by SCHORSCH, *Motor data sheet*.
[5] Electrical data provided by JEUMONT Electric, *Generator data sheet*.

Multiplatform Measurement Test Bench for Servodrives

[1]*Michal PAJKOŠ (1st year)*, [2]*Viktor ŠLAPÁK (1st year)*
Supervisor: [3]*František ĎUROVSKÝ*

[1,2,3]Dept. of Electrotechnics and Mechatronics, FEI TU of Košice, Slovak Republic

[1] michal.pajkos@tuke.sk, [2]viktor.slapak@tuke.sk, [3]frantisek.durovsky@tuke.sk

***Abstract*— This paper presents measurement test bench for servodrives. Hardware configuration is described as well as communication between used devices. Presented test bench can be used for static and dynamic measurements of servodrives as well as for precision and transmission error evaluation.**

***Keywords*—servodrive, test bench, Labview, TCP/IP, dynamic emulation of mechanical loads**

I. INTRODUCTION

Modern production and controlling processes increase their requirements for material flow control as well as need of precise position control of production support plants such as industrial robots. With a progress in all technological segments, the quality and accuracy of position sensors, electrical motors and precise transmissions, used not only in robotics, come to foreground.

In the present, servodrive, electrical motor with position sensor and precise transmission, is widely used in variety of applications. In some of them, power converter with motor control is included.

Whereas production of slow speed servomotors with high torque and compact size is technologically very difficult, servodrives with precise transmission is good alternative, especially in robotics. Permanent magnet synchronous motor is prevalent in this type of servodrives.

However, every transmission that is used in actuators shows nonlinearity, which causes angular transmission error. Inaccuracy of position sensing and control together with this angular transmission error determine angular transmission error of whole servodrive and thus its ability to reach its designated position.

In this paper we present different possibilities of multiplatform measurement stand based on Siemens and National Instruments technology. These measurements are performed on servodrives with cycloid gears as a representative of high precision gears.

II. MULTIPLATFORM MEASURING TEST BENCH

Multiplatform measuring test bench was developed to test and measure different types of electrical motors and servodrives. Measurements are focused on angular transmission error, repeated position accuracy, mechanical loading of actuators and emulation of loading torque etc. This multiplatform measuring system consists of control and measurement part.

Power converters, their control unit and other supported devices for both, loaded and loading servodrives, are built on Siemens components fig.1. Communication between loading and loaded part of the test bench is secured trough Profibus. As master control unit, SIMOTION D445-1 is used. This control unit includes also integrated PLC and it is able to manage up to 16 drives. Main advantage of this control unit is that it allows creating and testing various regulation structures for systems with variable moment of inertia.

There are also three addition cards for position sensor evaluation, which can evaluate signals from resolver, incremental sensor, EnDat and SSL, etc.

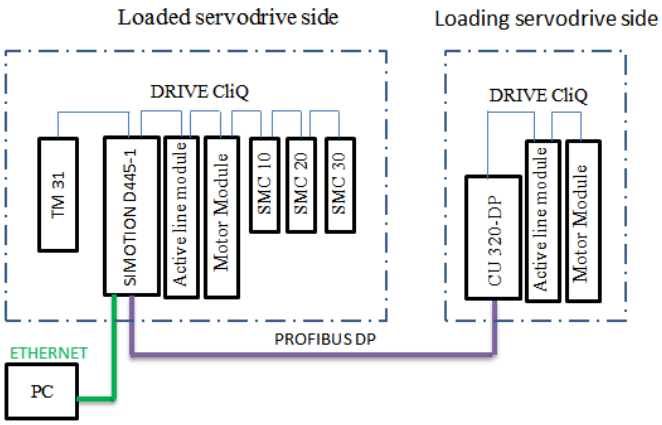


Fig. 1 Hardware configuration of control part of test bench

National Instruments CompactRIO (cRIO) represents measurement part of test bench. Main task of cRIO is to collect and process measured values of servodrives output position (from high resolution optical incremental position sensor by Renishaw) and torque as well as to receive data from SIMOTION D445-1.

CompactRIO and SIOMOTION communicate over Ethernet network using TCP/IP protocol with SIMOTION as server and cRIO as client fig.2. A Labview running PC is another TCP/IP server connected to the network as shown in fig. Its task is to control cRIO and to receive, process and store data sent from cRIO. Data are stored in text file on PC hard-drive.

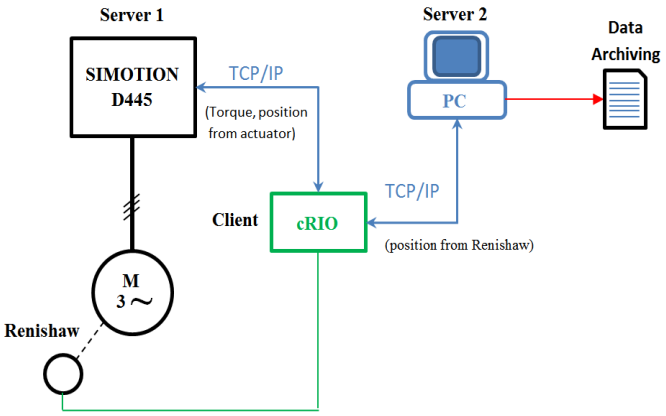


Fig. 2 Communication scheme of measuring stand

III. COMMUNICATION

Communication via TCP/IP protocol assumes specifying at least one server and one client. These specifications are required before connection is established and then data exchange is possible in both directions. Therefore communication is divided into three phases:

1.) Connection establishment
2.) Data exchange
3.) Connection closing

According to above statements the communication of multiplatform test bench devices is created.

In this hierarchy SIMOTION was programed as server, because of the fact that it controls the measuring test bench. Program for TCP/IP communication for SIMOTION part was created in LAD program language using special program blocks, and it was designed to send data, containing actual position and torque of measured actuator.

Client part of the test bench (represented by cRIO) is controlled and programed in Labview. Therefore this programing environment is used to create TCP/IP connection on client side [1].

Actual program for cRIO is shown in fig.3. Program consists from subsections, which are executed consequentially. Connection establishment part of fig. 3 consists from calling server SIMOTION as well as with server PC after which cRIO can acquire data about position and torque of the controlled servodrive. Data evaluation compares received data with data acquired through external sensors and adds timestamp to every sample collected. However, it is not recommended to store these data into cRIO flash memory due to memory size and therefore they have to be sent to the server PC. After measurement is finished TCP/IP is automatically closed.

Process of archiving data is running on the PC shown in fig.4. This program is as well created in Labview and it allows user to access and further analyze measured data, e.g. in Matlab.

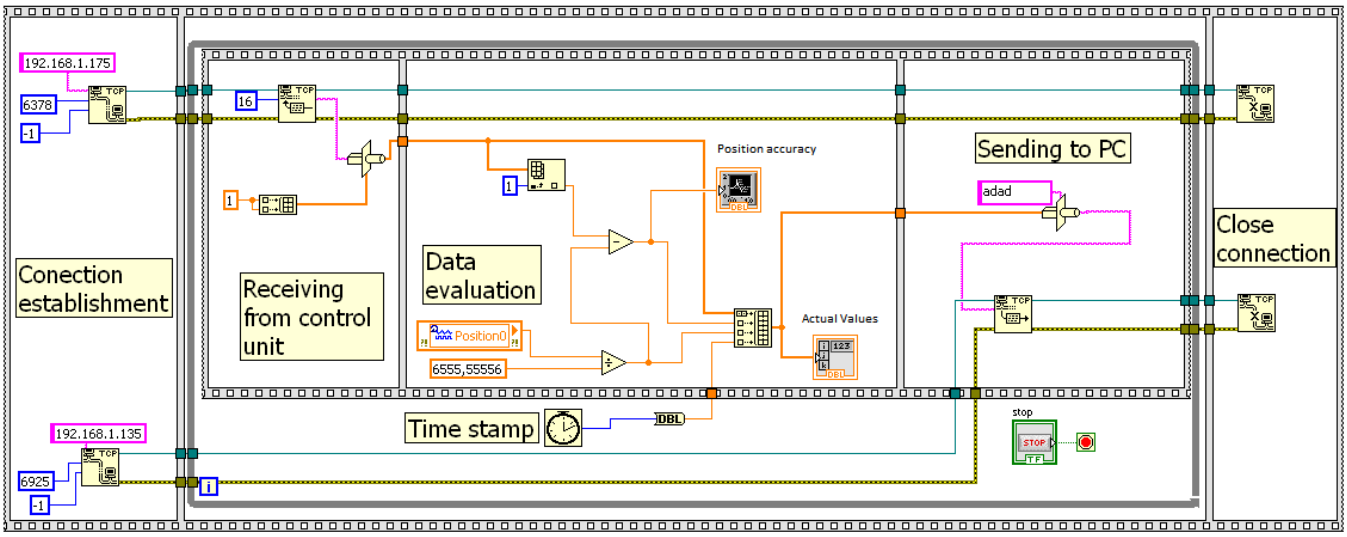


Fig. 3 CompactRIO program for TCP/IP communication

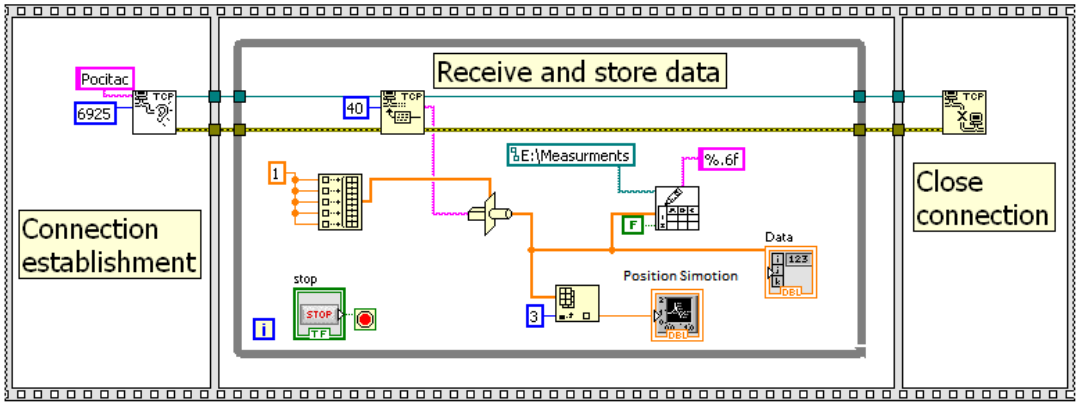


Fig. 4 Program for data logging in PC

IV. MEASUREMENTS TO BE REALIZED

As it was introduced in Chapter II, test bench was developed for the several types of measurements. These are following:

a) angular transmission error
b) static load testing
c) dynamic load testing
d) repeated position accuracy

In the first case, the partial results was published in [2]. Last type of measurement has not been performed yet. Concerning static and dynamic load testing, so called methods for dynamic emulation of mechanical loads (DEML) [7] was realized in order to research on advanced dynamometer control strategies. Different approaches for DEML have been developed in last decade. Authors in [3]-[7] use a different approaches for the dynamometer control. The aim is to control the dynamometer (in our case it is loading actuator side) in such way that it behaves like real mechanical load connected to the drive under test. In [7] authors successfully realized emulation of several mechanical loads as two-mass flexible system or simple pendulum.

One of the difficulties with realization in DEML is signal delay caused by communication buses. Authors in [7] used an experimental setup where signal delays have significant influence on dynamometer performance. Example of signal delay is depicted in Fig. 5. Blue signal is sent from RT-Lab to Siemens Simoreg converter and the same signal is received back after the time delay as green signal. This time delay behaviour can be described as a transport delay nonlinearity and it is caused by communication bus and sampling time of converter I/O and data acquisition card circuits. It can be compensated either by improved control structure or by using different hardware. However, the control algorithm should be kept as simple as possible and focused on emulation structure itself. Therefore, better solution is to avoid the signal delay by using hardware with improved performance and thus, smaller transport delays. The test bench introduced in this paper was built to satisfy these requirements. Presented hardware was chosen in order to reach the best performance with industrial components, e.g. communication bus DriveCliq is highly-dynamic interface and signals for electronics are available in rating of µs.

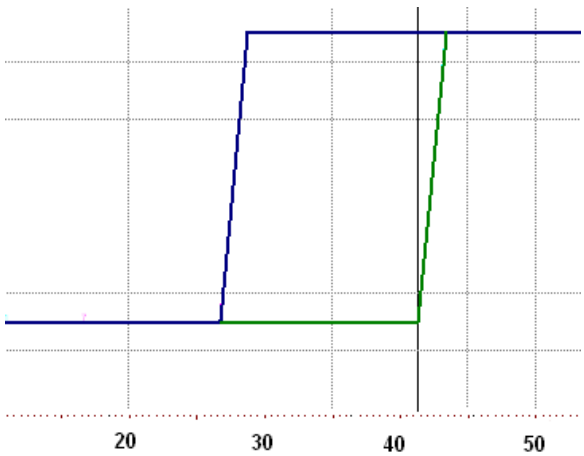


Fig. 5 Example of signal delay bus measurement in Matlab – Simoreg multiplatform test bench, x-axis [time/ms], y-axis [signal value]

V. CONCLUSION

Multiplatform test bench is used for evaluation of precision and performance of high precision servodrives which are widely used in present. Important aspect of this test bench is communication between used systems based on TCP/IP which is necessary for data evaluation. Measurements preformed on this test bench are all based on predefined methodic. Moreover presented multiplatform test bench is suitable for research of advanced dynamometer control strategies based on DEML.

REFERENCES

[1] Basic TCP/IP Communication in LabVIEW, aviable at: <http://www.ni.com/white-paper/2710/en>.
[2] Hric, M., Ďurovský, F., Fedák, V.,: "Vplyv nelinearít cykloidnej prevodvky na presnosť polohovania", ATP journal, 09/2012 pp. 42 – 44
[3] Žalman, M., Macko, R.; „Design and realization of programmable emulator of mechanical loads," *16th IFAC World Congress*, Volume 16, Czech Republic, July 2005
[4] Arellano-Padilla, J.; Asher, G.M.; Sumner, M.; „Control of an AC dynamometer for Dynamic Emulation of Mechanical Loads With Stiff and Flexible Shafts," *Industrial Electronics, IEEE Transactions on*, vol.53, no.4, pp.1250-1260, June 2006
[5] Rodic, M.; Jezernik, K.; Trlep, M.; "Mechatronic Systems' Control Design Using Dynamic emulation of Mechanical Loads," *Automatika: Journal for Control, Measurement, Electronics, Computing and Communications*, vol. 47, no 1-2, pp. 11-18, ISSN 0005–1144, May 2006
[6] Karol Kyslan, František Ďurovský, Control of a Test Bench for Dynamic Emulation of Mechanical Loads, Procedia Engineering, Volume 48, 2012, Pages 352-357, ISSN 1877-7058
[7] K. Kyslan, "*Load Torque Emulator*," Dissertation thesis, Technical University of Košice, FEEI, 2012, (in Slovak).

Optimal placement and sizing of Static Var Compensator with using genetic algorithm

[1]*Roman JAKUBČÁK (2st year),* [2]*Miroslav KMEC (1st year)*
Supervisor: [3]*Ľubomír BEŇA*

[1,2,3]Dept. of Electrical Power Engineering, FEI TU of Košice, Slovak Republic

[1]roman.jakubcak@tuke.sk, [2]miroslav.kmec@tuke.sk, [3]lubomir.bena@tuke.sk

***Abstract*— This article discusses about FACTS (Flexible Alternating Current Transmission System) devices in power systems. The article is focusing mainly on decreasing active power losses in power system. Minimum active power losses in power system can be achieved by appropriately placed SVC (Static Var Compensator) with optimal set of parameters. All simulations were performed in program Matlab with using genetic algorithm toolbox.**

***Keywords* — power system, SVC, genetic algorithm.**

I. INTRODUCTION

Nowadays, the requirements for possible ways to control the power system increase rapidly, mainly due to the constant rise of electricity demand. This trend, together with market liberalization causes problems in the management of power systems and security of power system.

With reactive power control we are able increase stability of power system and at the same time reduce active power losses occurring in the transmission of power. For this purpose, you can use FACTS devices. One of these devices is the SVC, which allows you to manage reactive power at the place of his connection into power system.

II. STATIC VAR COMPENSATOR

Static Var Compensator (SVC) is a general term used either for a thyristor controlled, thyristor switched reactor, thyristor switched capacitor or some of them combined. SVC includes separate equipment for leading and lagging vars, the thyristor controlled or thyristor switched reactor for absorbing reactive power and thyrirstor switched capacitor for supplying the reactive power [1].

Thyristor Controlled Reactor (TCR) – a shunt connected, thyristor controlled inductor whose effective reactance is varied in a continuous manner by partial conduction control of the thyristor valve. Conduction time is controlled by a thyristor based AC switch with firing angle control [3].

Thyristor Switched Reactor (TSR) – a shunt connected, thyristor switched inductor whose effective reactance is varied in a stepwise manner by full or zero conduction operation of the thyristor valve. *TSR* is made up of several shunt connected inductor which are switched in and out by thyristor switches without any firing angle controls in order to achieve the required step changes in the reactive power consumed from the system.

Thyristor Switched Capacitor (TSC) – a shunt connected, thyristor switched capacitor whose effective reactance is varied in a stepwise manner by full or zero conduction operation of the thyristor valve [1].

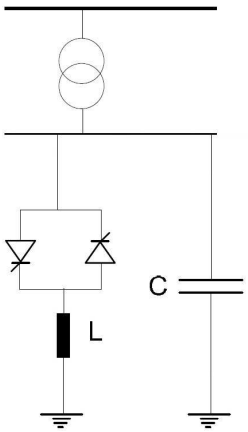


Fig. 1. Static Var Compensator

Installing an SVC at one or more suitable points in the network we can increase transfer capability and reduce losses while maintaining a smooth voltage profile under different conditions and mitigate active power oscillations [5].

III. SVC MODEL

The SVC may works in two modes. In inductive mode device absorbs reactive power. In capacitive mode device provide reactive power. SVC model consist of variable reactive power source connected into the node. It was considered that the maximum possible consumed and supplied reactive power was 50 MVAr.

$$-50 \le Q_{SVC} \le 50 \tag{1}$$

IV. GENETIC ALGORITHM

The genetic algorithm (GA) is a method for solving optimization problems that is based on natural selection. The GA repeatedly modifies a population of individual solutions.

At each step, the GA selects individuals at random from the current population to be parents and uses them produce the children for the next generation. Over successive generation, the population evolves toward an optimal solution.

The GA is suitable to solve a variety of optimization problems that are not well suited for standard optimization algorithms, including problems in which the objective function is discontinuous, nondifferentiable, stochastic, or highly nonlinear [2].

The GA uses three main types of rules at each step to create the next generation from the current population:

Selection rules – select the individuals, called parents that contribute to the population at the next generation.

Crossover rules – combine two parents to form children for the next generation.

Mutation rules – apply random changes to individual parents to form children [2].

V. Problem Formulation

The main goal is to determine the best location for SVC and its value by minimizing the objective function. For this purpose we use genetic algorithm. In this work we are following one goal and this is to reduce active power losses. Objective function includes equations for determining the total active power losses in the power system and penalty function. This function has a value of zero if all constraints are satisfied. If it does not the value is ten.

The main constraints in the optimization process are [4]:

1. Voltage limits
2. Max/min reactive power generated by generator
3. Max/min reactive power generated or consumed by SVC
4. Max/min position on transformer tap changer
5. Max permissible current flow through the line

Optimization without SVC – The first optimization process does not consider the use SVC. In this case the controlled variables are:

- Generator bus voltage magnitude
- Transformer tap ratios

Generator voltages and transformer tap ratios are used as control variables. Transformer tap ratios are treated as continuous variables during the optimization, after which they are adjusted to the nearest physical tap position.

The possible values of reactive powers supplied by generators in nodes 3 and 5 are in the range:

$$0 \le Q_{gen\,3} \le 100\,[MVAr\,] \tag{2}$$

$$100 \le Q_{gen\,5} \le 300\,[MVAr\,] \tag{3}$$

Optimization with SVC – In this case we consider the use of SVC. We added into optimization process two control variables. They describe location and value of SVC.

In all cases we consider same stop criterion. The simulation stops when it reaches the maximum number of iterations.

VI. Simulation Results

We consider 14 nodes power system. The first five nodes are on 400 kV level, others are on 110 kV level. The following figures and tables show a simulated network, generations and loads in nodes, voltage profile and total active power losses before optimization and after optimization with and without SVC.

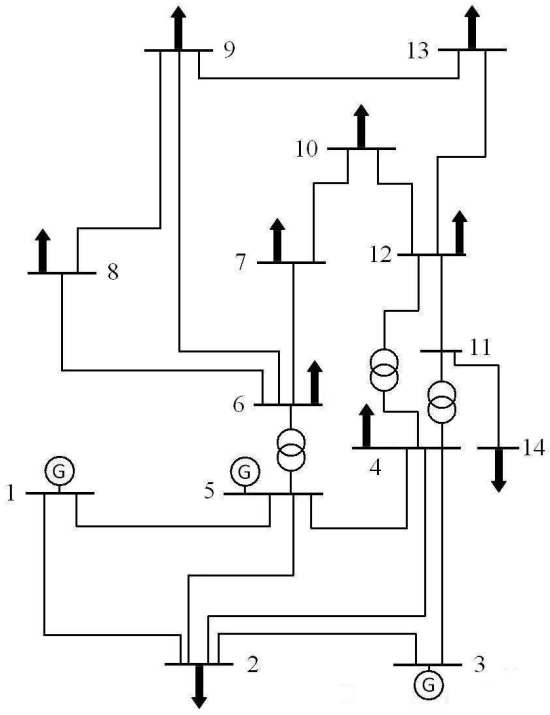


Fig. 2. 14 nodes power system

TABLE I
Generations and Loads in Nodes

Node	Before optimization		Without SVC		With SVC	
	P gen [MW]	Q gen [MVAr]	P gen [MW]	Q gen [MVAr]	P gen [MW]	Q gen [MVAr]
1	639	-124	636	-173	636	-211
2	-400	-175	-400	-175	-400	-175
3	100	50	100	66	100	62
4	-400	-175	-400	-175	-400	-175
5	400	200	400	225	400	220
6	-40	-15	-40	-15	-40	-15
7	-40	-15	-40	-15	-40	-15
8	-40	-15	-40	-15	-40	-15
9	-40	-15	-40	-15	-40	-15+43.6
10	-40	-15	-40	-15	-40	-15
11	0	0	0	0	0	0
12	-40	-15	-40	-15	-40	-15
13	-40	-15	-40	-15	-40	-15
14	-40	-15	-40	-15	-40	-15

Positive values in the table represent generation, negative represent consumption active/reactive power.

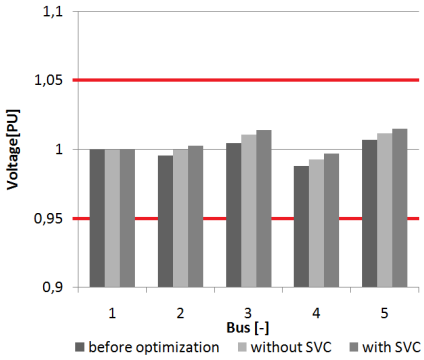


Fig. 3. Voltage profile on 400 kV level

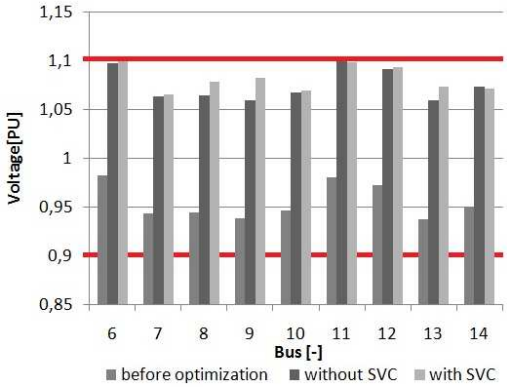


Fig. 4. Voltage profile on 110 kV level

In second case, optimization process determined that the best location for placement SVC is node 9 with value 43.6 MVAr. After connecting SVC into node 9 total active power losses in power system decrease from 16.42 MW to 15.9 MW. It should be noted that total active power losses before optimization were 18.61 MW. Changes in the active power losses in the lines are shown in fig. 5. It should be noted that after connecting SVC into node 9 active power losses increase on 400 kV level from 6.69 MW to 6.74 MW. But on 110 kV level decrease from value from 9.73 MW to 9.16 MW.

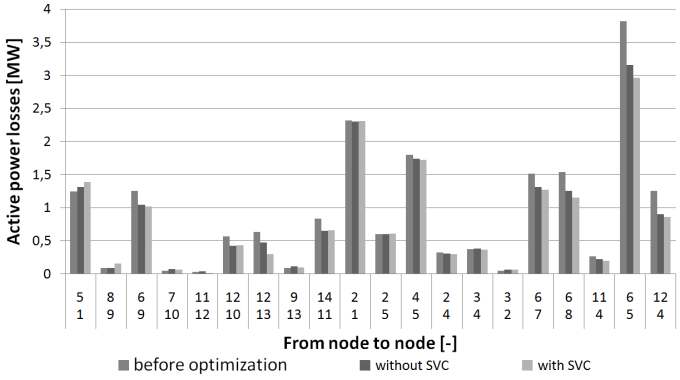


Fig. 5. Active power losses in all lines

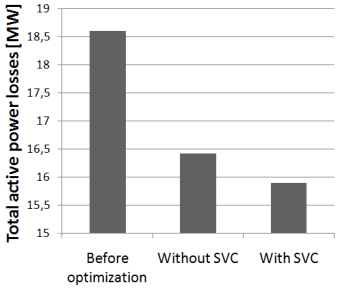


Fig. 6. Total active power losses

VII. Conclusion

As shown in this article using SVC, we are able to control voltage at the node to which SVC is connected and reduce active power losses in power system simultaneously. Because financial costs of these devices are high, we should use them to achieve multiple purposes simultaneously. It is all on an individual suitability of the location of SVC in power system.

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 contract No. 1/0166/10 and by Slovak Research and Development Agency under the contract No. APVV-0385-07 and No. SK-BG-0010-08.

References

[1] HINGORANI, G. N., GYUGYI, L., “Understanding FACTS. Concepts and technology of Flexible AC transmission Systems,” New York: IEEE Press, 2000, 432 pp, ISBN 0-7803-3455-8

[2] Genetic algorithm and direct search toolbox, Available on internet: <http://www.mathworks.com/help/releases/R13sp2/pdf_doc/gads/gads_tb.pdf>

[3] JOHNS, A. T., TER-GAZARIAN, A., WARNE, F.: “Flexible ac transmission systems (FACTS),” The Institution of Electrical Engineeres, 1999, 592 pp, ISBN 0-85296-771-3

[4] MOMOH, J.: Electic Power System Application of Optimalization Hovard University Washington, D.C., 2001. 478 s. ISBN 0-8247-9105-3.

[5] Static Var Compensator, Available on internet: <http://www.abb.com/cawp/gad02181/c1256d71001e0037c1256bd60043b75c.aspx. >

Phonetically Balanced Slovak Speech Corpus for Text-To-Speech Synthesis

[1]*Martin SULÍR (1st year)*
Supervisor: [2]Jozef JUHÁR

[1,2] Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

[1]martin.sulir@tuke.sk, [2]jozef.juhar@tuke.sk

***Abstract*—This paper describes the development of a phonetically balanced Slovak speech corpus for text-to-speech synthesis. HMM-based speech synthesis is a corpus-based approach to speech synthesis that generates speech parameters from statistical models built on recorded speech from a large database, so selecting optimum text from available text data with balanced phonetic elements distribution is an important task for developing high quality TTS system based on HMM speech synthesis method. We describe our text selection technique based on Greedy selection algorithm, the results of selection and the phonetically balanced corpus recording and processing.**

***Keywords*— greedy algorithm, speech synthesis, speech corpus, text selection**

I. Introduction

With the development of computer technology, the interaction between human and computer based on the automatic speech recognition, speech synthesis and natural language understanding has become a challenge in the world. Text to speech (TTS) systems represents one of the most important part of the speech interaction with computer and the research in this area is carried out in many countries and companies around the world. Speech synthesis nowadays represents particularly computer systems that can convert input text to speech [1]. The main task of these systems is make life easier, either to people with physical disabilities such as the blind people or to totally ordinary people who use these systems to facilitate day to day operations. Research in this area is aimed to the point, when it will be possible to use these voices in various spheres of life, without that they were somewhat limited and acts unnatural.

The basic requirements which should be met in every TTS system are the highest possible intelligibility, fluency and naturalness of speech at the output of this system. The quality of this output is defined by the quality of speech corpus in the case of HMM speech synthesis [2] (which represents our field of research). Most of the corpus-based TTS systems, like HMM speech synthesis or unit selection synthesis, require a speech corpus with phonetically balanced sentences that should have high coverage of phonetic contextual units to provide high quality in speech synthesis system, so the speech corpus derived from the mother text is a basic and important part in the corpus-based speech synthesis.

The principle of corpus construction is to use the least material to cover the most natural phenomena and for this selection greedy algorithm was used [3]. This paper describes creation of the new Slovak male and female speech corpus for HMM speech synthesis.

II. Greedy selection

The greedy algorithm is a simple iterative technique for constructing a subset of sentences from a large set of sentences to cover the largest unit space with the smallest number of sentences [4][5]. This is achieved in a first step by choosing a unit size by which to define linguistic unit coverage. In this paper, the base unit originally chosen was a diphone. The criteria for whether a sentence gets included in the subset depends on how phonetically varied the sentence is, given by the distribution of phones and diphones within the sentence.

The greedy optimal selection process consists of assigning specific score to every sentence according to the number of out of cover units and in cover units in the sentence. At each iteration, the algorithm picks the most useful sentence to include in the subset of sentences, removes the processed sentence from the large set and updates the target cover set and the sentence score in the large set. Iterating these steps until a termination criterion is reached, the subset is constructed.

The following equation was proposed to select the phonetically balanced sentences [6]:

$$S = \sum_{i=1}^{I} \left| r_i - \frac{n_i + n'_i}{n} \right| \tag{1}$$

where,

$$n = \sum_{i=1}^{I} (n_i + n'_i) \tag{2}$$

and I is a number of different units to be included in the selected sentences, n_i is the number of occurrences of the ith unit in the selected sentences, n'_i is the number of occurrences of the ith unit in the currently processed sentence and r_i is the desired relative frequency of the ith unit.

The optimal desired value of score, at which selection algorithm should be terminated, is equal to zero. However, it

is very difficult to achieve this ideal value in the real conditions the algorithm is usually terminated when it reaches the desired number of sentences.

III. Methodology of Text Selection Procedure

The methodology of the text selection procedure is presented in this section. The aim of this procedure is to obtain the Slovak phonetically balanced text corpus, which forms a basis for the subsequent recording of the speech corpus for the HMM-based speech synthesis in Slovak language. The system diagram of the whole development process is shown in the Figure 1.

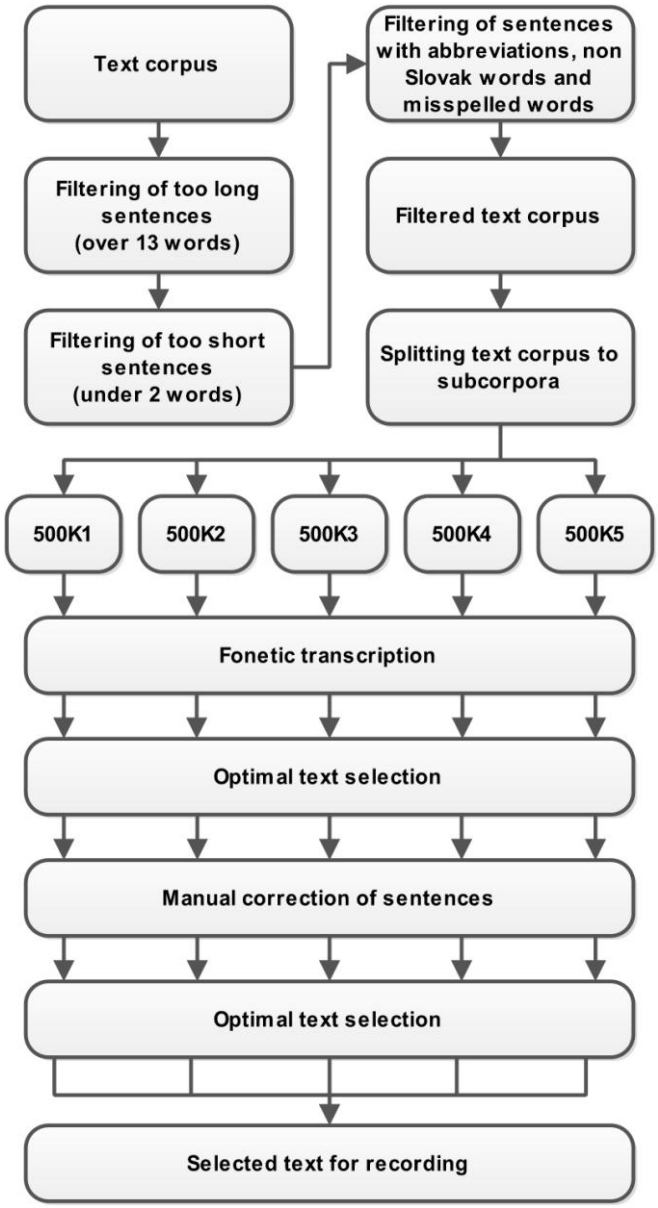


Fig. 1. The system diagram of the text selection procedure.

As we can see, the first block in this diagram is a text corpus. This input text corpus represents text data, which were collected for the purposes of language modeling in speech recognition system. The following data represents the Slovak sentences (over 10 million sentences) which are divided into separate lines and normalized. Therefore, these text data constitute a sufficient basis for building the phonetically balanced text corpus for speech synthesis, but it was necessary to further modify them.

The first modification that had to be done was filtering of too long sentences because these sentences would be too difficult to record and in many cases they contain meaningless concatenated words and normalization tags. The same procedure was carried out with too short sentences that contain only one word.

The next step was filtering of sentences that contain the words whose pronunciation is not clearly defined in the Slovak language. In many cases it was the abbreviations, the non-Slovak words and the misspelled words. For this filtering regular expressions were used.

The output of these steps was the text corpus, which contains two and a half million sentences, so something more than 7 million of sentences were filtered out. Because there was a limit of processing power, it was not possible to process such an amount of sentences using the algorithm for optimal text selection. The filtered text corpus was divided into five subcorpora, where each subcorpus contains approximately five hundred thousand of sentences and each of them were then processed separately. The separate processing of smaller subcorpora enables us to use the standalone or combined corpora in the speech synthesis for evaluation its quality depending on the size of the speech corpus.

Each subcorpus was subsequently transferred to its phonetic form, because the optimal text selection is carried out at the level of phonemes. The optimal text selection algorithm was then applied on each transcribed subcorpus. This algorithm selected approximately 1000 sentences from the input subcorpus which represented the phonetically balanced subcorpus. After this selection, it was necessary to manually check the individual selections and delete the sentences that still contain the inappropriate words for recording. The last step of the subcorpora processing was re-selection of the most appropriate sentences from the manually checked selected sentences of each subcorpora. The resulting phonetically balanced corpus was obtained by combining the subcorpora together.

IV. Optimal Text Selection Results

The results of optimal text selection with the proposed methodology are shown in the Table I.

TABLE I
Optimal Text Selection Results

	Number of different phonemes	Number of selected sentences	Number of different diphones	Total number of diphones in selection
500K1	45	962	1211	40 919
500K2	45	933	1195	39 229
500K3	45	889	1193	38 542
500K4	45	884	1201	37 749
500K5	45	858	1198	36 610
Total	**45**	**4526**	**1260**	**193 049**
Percentual coverage:			1260/1600 = 78,75 %	

As we can see, the final version of corpus contains 4526 sentences and the total number of different diphones is 1260. Considering that the speech synthesis in Slovak language uses from 1200 to 1600 diphones [7], so we can say that the percentual coverage of this selection is in the worst case (If we consider 1600 diphones) 78,75%. The following percentage coverage of the elements of language is a good result

regardless of the complexity of the Slovak language as such. The frequency of diphones in the selected text is shown in the Figure 2.

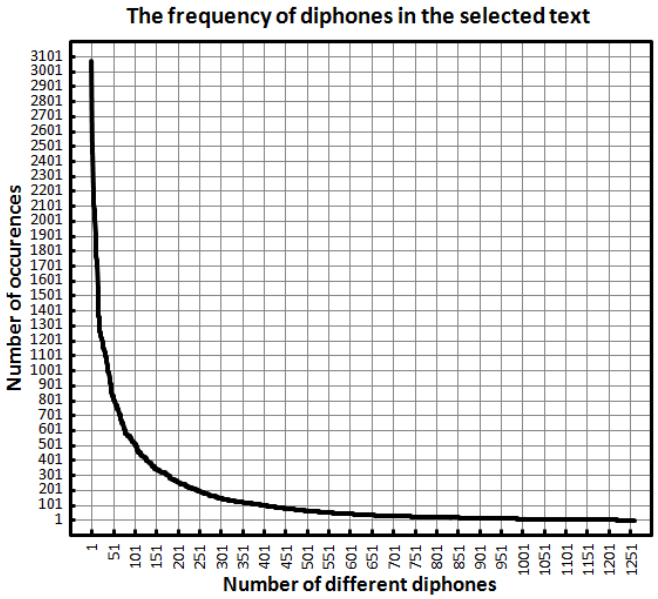


Fig. 2. The frequency of diphones in the selected text

As we can see, the most numerous diphone, which is located in the selected text, has a higher frequency than 3000 occurrences and more than four hundred diphones are represented by more than one hundred occurrences, so that the distribution of diphones in the selected text can be considered sufficient. The detailed frequency of the first twenty diphones is shown in the following table.

TABLE II
TOP 20 DIPHONES IN SELECTION

No.	Diphone	Number of occurrences	No.	Diphone	Number of occurrences
1.	-n-a-	3077	11.	-r-e-	1804
2.	-t-o-	2482	12.	-d-o-	1767
3.	-p-o-	2461	13.	-r-o-	1733
4.	-p-r-	2207	14.	-r-a-	1700
5.	-s-a-	2188	15.	-o-m-	1544
6.	-J-e-	2115	16.	-v-a-	1457
7.	-o-v-	2072	17.	-l-a-	1365
8.	-k-o-	2016	18.	-v-e-	1359
9.	-j-e-	2006	19.	-h-o-	1304
10.	-s-t-	1908	20.	-o-r-	1265

V. SPEECH CORPUS RECORDING AND PROCESSING

The final step in the creation of the phonetically balanced speech corpus was recording and processing of the selected sentences. The recording took place in a professional recording studio with one male and one female speaker. The same set of 4526 phonetically balanced sentences was recorded by each of them. A Neumann TLM 103 cardioid condenser microphone with SPL Gold Mike Mk2 pre-amplifier and a hard disk recording system equipped with RME Fireface 400 audio interface was used during the sessions. A 48 kHz sampling frequency and 16 bit resolution were used.

Because the recording was carried out in parts, where each part contains one hundred sentences, it was necessary to process them into isolated sentences. Also, the segments of silence had to be aligned at about 200 milliseconds at the beginning and end of the sentences [8]. Each of the speech corpora, the male and also the female, consists approximately of more than five hours of pure speech (without segments of silence), so we can say that each subcorpus contains more than one hour of pure speech.

VI. CONCLUSION

In this paper we presented the development of the phonetically balanced Slovak speech corpus for text-to-speech synthesis in Slovak language. This speech corpus contains 4526 sentences with 78.75% diphone coverage, so this coverage of the Slovak language can be considered sufficient.

The big advantage of our subcorpus approach is that it is possible to split the final corpus to the five subcorpora and we can examine the impact of corpus size on the quality of speech synthesis.

Our future research direction is to construct a speech database with our obtained speech corpus, and we will realize a speech synthesis system based on HMM speech synthesis technique with this new male and female database.

ACKNOWLEDGMENT

The research presented in this paper was supported by the Ministry of Education, Science, Research and Sport of the Slovak Republic under research projects No.3928/2010-11 (50%) and by Research & Development Operational Program funded by the ERDF under the ITMS project 26220220155 (50%).

REFERENCES

[1] P. Taylor, *Text-to-Speech Synthesis*. University of Cambridge, Cambridge: Cambridge University Press, 2009. p. 597. ISBN 978-0-521-89927-7

[2] H. Zen, T. Nose, J. Yamagishi, S. Sako, T. Masuko, A. Black, K. Tokuda, *The HMM-based Speech Synthesis System (HTS) Version 2.0*. *HMM-based Speech Synthesis System*, 2007.

[3] J. P. H. van Santen, A. L. Buchsbaum, *Methods for Optimal Text Selection*. In Proceedings of the Eurospeech, 1997. pp. 553-556.

[4] S.M. Habib, F. Alam, R. Sultana, S. A. Chowdhur , M. Khan, *Phonetically balanced Bangla speech corpus*. In Proceedings of Conference on Human Language Technology for Development, Egypt 2011. pp. 87 – 94

[5] B. Bozkurt, O. Ozturk, T. Dutoit, *Text design for TTS speech corpus building using a modified greedy selection*. In Proceedings of the Eurospeech, 2003. pp. 277 – 280

[6] J. Psutka, L. Müller, J. Matoušek, V. Radová, *Mluvíme s počítačem česky*. Prague: Academia, 2006. p. 752, ISBN 80-200-1309-1

[7] J. Juhár, *Rečové technológie v telekomunikačných a informačných systémoch*. Košice : Equilibria, 2011. p. 519, ISBN 978-80-89284-75-7

[8] J. Kominek, A. Black, *CMU ARCTIC databases for speech synthesis.* Carnegie Mellon University, 2003.

Rolling Speed and Torque Prediction in Wire Rod Mill

[1]*Tomáš BOROVSKÝ (2st year)*
Supervisor: [2]František ĎUROVSKÝ

[1,2]Dept. of Electrical Engineering and Mechatronics, FEI TU of Košice, Slovak Republic

[1]tomas.borovsky@student.tuke.sk, [2]frantisek.durovsky@tuke.sk

***Abstract*—The paper presents and analyses methods of rolling speed and torque calculations on wire rod rolling. These rolling parameters are one of the most important points in mathematical model of wire rod rolling mill. Hence, accuracy in their computation in wide range of input data has crucial effect on capability of the mathematical model. The mathematical model will be used to analyze and tuning the control structures. Results are compared with experimental data obtained from real rolling processes on wire rod mill in Slovakia Steel Mills, Strážske.**

***Keywords*—wire rod and bar rolling, hot rolling, rolling torque**

I. INTRODUCTION

Control structures on rod and bar rolling mill operates in wide range of speed, because rolling speed varies from *0.1m/s* on the first stand up to *120m/s* on the last stand. Because of a problematic technique of material real speed measurement, speed control should work with theoretically calculated speed, so deviations from real speed occurs usually. Moreover, speed control should be robust enough, to overcome problems of any changes in rolling parameters, which have a big effect to real speed of material. Normally, control system is tuned on the site during the rolling. But it takes a production time and brings undesirable risk. With mathematical model of rolling mill, existing control structure can be tuned and new control strategy can be tested without any influence on production. Model of workpiece, where rolling force is output parameter, should be compared with measurement results to verify its capability. Two methods of rolling force calculation are presented and compared with experimentally obtained data in this paper.

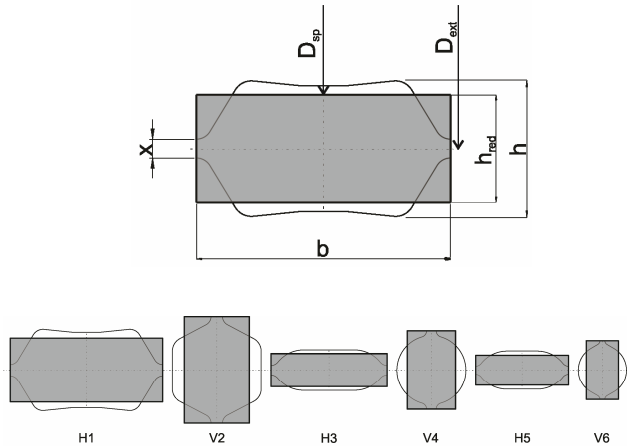


Fig. 1. Groove geometry and pass schedule used for calculations

II. EFFECTIVE ROLL DIAMETER AND ROLLING SPEED

Calculating rolling speed through rolling stands is one of the most important tasks involved in roll pass design, because it is basis in computation of mass balance and other rolling parameters too. In flat rolling processes, the rolling speed at a given pass can be calculated much easier, because the roll radius is constant along the roll axis direction. This condition is not fulfilled in rod and bar rolling, because rolls with grooves are used. In order to overcome this difficulty, the term mean roll diameter D_{sp} has been introduced. There have been several models that determine mean roll radius in rod and bar rolling. In this paper, Wusatowski model has been used to calculate mean roll diameter, and it is given by:

$$D_{sp} = D_{Ext} - \left(\frac{S}{b} - x \right) \tag{1}$$

where S is the exit cross-sectional area of deformed workpiece at given pass, b is maximum width of deformed workpiece, x is gap between the rolls and D_{ext} is maximum diameter of rolls [1].

It is assumed, that the speed of the workpiece at neutral point is equal to circumferential speed of roll at mean roll diameter. Hence we can calculate the output speed v_m of the workpiece by (2), where v_{m0} is incoming workpiece speed, S_0 is cross sectional area of incoming workpiece and S_m is cross sectional area of workpiece in neutral poins. For this calculation is necessary to know neutral angle γ. Neutral angle for low carbon steel rolling with speed from *0.2* to *1m/s* is given by (2) and (3), where α is impact angle, γ is friction angle, T is workpiece temperature and $a=0.8$ is factor of rolls effect [1].

$$v_m = \frac{v_{m0} S_0}{S_m}; \ \gamma = \frac{\alpha}{2}\left(1 - \frac{\alpha}{2\rho}\right) \tag{2}$$

$$\mu = a(1{,}05 - 0{,}0005T) \tag{3}$$

III. CALCULATION OF ROLLING TORQUE

In continuous hot rod and bar rolling processes, the roll

force predicting is one of the most important tasks for process designers in designing and optimizing a rolling schedule. Usually, finite element method is used to compute it. But, if computational time and vague boundary rolling conditions are

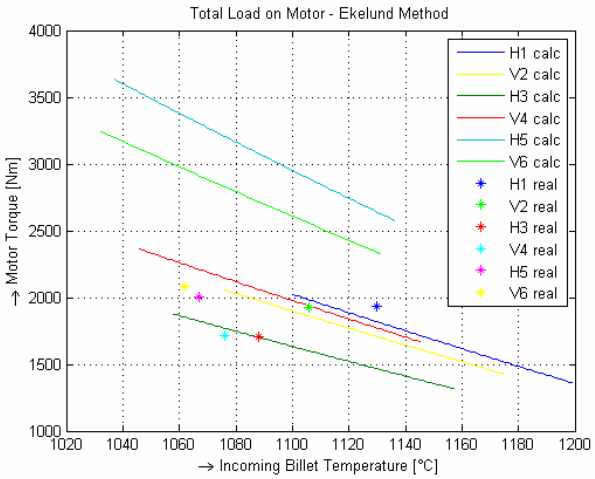


Fig. 3. Comparison of total motor torque calculated using Ekelund method versus experimental obtained data

considered, there have been several attempts to develop an approximate model as an alternative [1]. Especially, for dynamic mathematical model of rolling mill should be used a

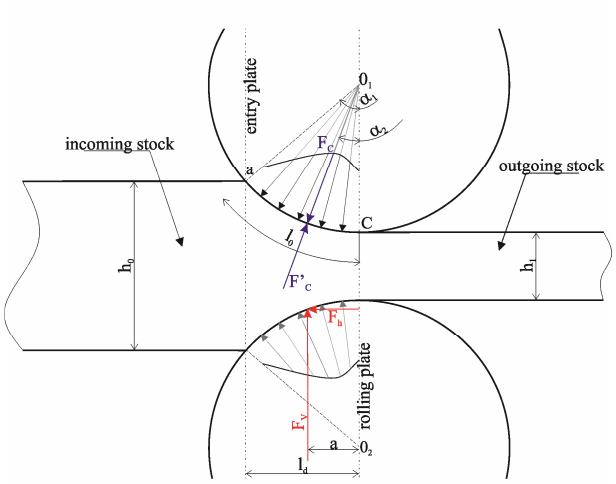


Fig. 2. Rolling force in rod and bar rolling

simple model. For these purposes is not necessary to guarantee high accuracy in calculations. More important is to get stable and satisfactory results in wide range of input parameters and short computational time. This paper deals with methods of rolling force and rolling torque computation. For these calculations, Ekelund method and Deformation energy method are used.

A. *Ekelund method*

Rolling force can be estimated from general formula (4), where p_{str} is average flow stress of the workpiece in the groove, S_d is projected contact area of the workpiece in the groove. If we consider, that the horizontal part of rolling torque is very small compared to its vertical part, total rolling force F_v is equal to its horizontal part. We can estimate the rolling torque per roll by assuming that the force F acts in the middle of the arc of contact, as is shown in Fig.2. Thus, rolling torque per roll is given by (4), where a is force arm and F_v is total rolling force [4]. Equation (4) is ideally for a frictionless situation. Thus, the higher the friction coefficient between rolls and the workpiece, the more the divergence [2].

$$F_v = p_{str} S_d \; ; \; M_{Fv} = F_v a \tag{4}$$

B. *Energy of deformation*

This method is usually used in calculations for rod and bar

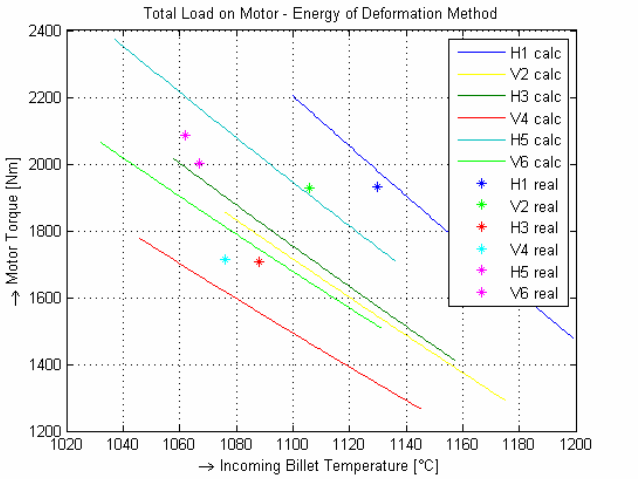


Fig. 4. Comparison of total motor torque calculated using Energy of Deformation Method versus experimental obtained data

rolling processes, because of difficult geometry of roll grooves and consequently problematic determination of rolling force F_v. Rolling torque is calculated by (5), where W_{def} is energy consumption and L is total length of input workpiece.

$$M_{Fv} = W_{def} \frac{D_{sp}}{2L} \; ; \; W_{def} = p_{str} V \ln(\gamma) \tag{5}$$

According Fink, Energy of Deformation can be estimated by (5), where V is volume of rolled workpiece and γ is elongation factor [3].

IV. Experimental Results

Since there is no sensor for rolling force measurement, this can be estimated from motor torque feedback only. But total motor torque does not consist only from the rolling torque. It is given by (6), where M_t is no load friction torque, M_{tah} is tension torque, M_{cap} is friction torque from rolls pins and M_v is rolling torque.

$$M_m = M_t + M_{tah} + M_{cap} + M_v \tag{6}$$

No load friction torque M_t can be obtained from motor torque feedback during no load operation. Friction torque from rolls pins is given by (7), where F_v is rolling force, f_{cap} is rolls pins friction factor, for considered conditions is estimated to *0.07*, and D_{cap} is rolls pins diameter [4].

$$M_{cap} = \frac{2F_v f_{cap} D_{cap}}{4} \tag{7}$$

$$M_{Tension} = M_{NoTension} - M_{WithTension} \tag{8}$$

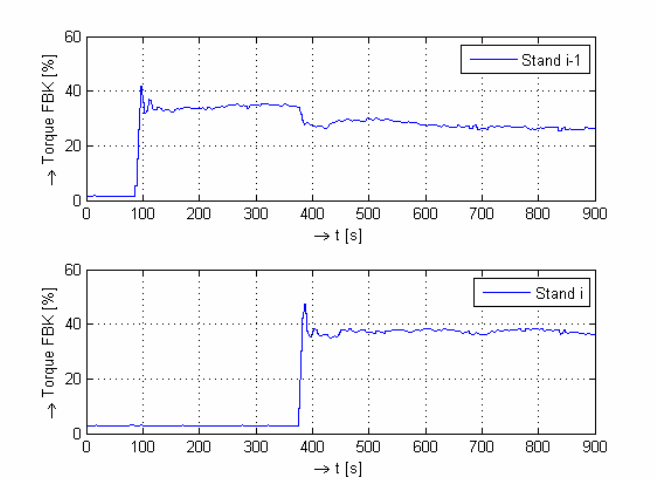


Fig. 5. Tension and no tension status of Stand i-1

Tension torque M_{Tenion} is torque which is related to tension between the stands. It can be estimated from experimental data, by (8). $M_{NoTenion}$ is motor torque of given stand during time interval when workpiece is already rolled in given stand, but not in stand after yet. $M_{WithTension}$ is motor torque of given stand during time interval when workpiece is rolled by given stand and by stand after. $M_{WithTension}$ is lower than $M_{NoTenion}$, because of tension from stand after, as is shown in Fig.5.

All these additional parts of total torque which load the motor are added to rolling torque and products are consequently compared with experimental data. Results are presented in Fig. 3 and Fig. 4 for Ekelund Method and Energy of Deformation Method, respectively.

V. Conclusion

Rolling torque and rolling speed calculation for groove rolling are presented in this paper. Two methods of rolling torque calculation are compared with experimental data. Results are presentet in Tab.1. Energy of deformation method brings more acceptable results, especially for last two stands. A big deviation in calculation using Ekelund method is probably due to the difficult groove geometry and consequently vague projected contact area estimation. This implies, that energy of deformation method is more suitable for groove rolling generally. Observed deviation of calculated torque using this method can be caused by inaccurate cross sectional area or temperature definition.

TABLE I
Difference in total torque

Stand	H1	V2	H3	V4	H5	V6
M_{DefEng} [%]	2.4	12.8	7.0	5.6	8.2	9.3
$M_{Ekelund}$ [%]	5.9	3.5	0.4	25.2	64.9	41.9

M_{DefEng} is difference in % of measured torque between measured torque and calculated torque using energy of deformation method. $M_{Ekelund}$ is difference in % of measured torque between measured torque and calculated torque using Ekelund method.

References

[1] Youngseog Lee, *Rod and Bar Rolling*. Pohang, Korea, POSCO Technical Research Laboratories, 2004
[2] NIIR Board of Consultants & Engineers, *The Complete Technology Book on Hot olling of Steel*, Delshi - India, National Institute of Industrial Research, 2010
[3] Karel Stýblo, *Strojní zažízení válcoven a tažíren: 1. část*, Ostrava, VŠB v Ostravě, 1969
[4] Mária Kollerová a Milan Žídek a Bohumil Počta a Vladimír Dědek. *Valcovanie*, Bratislava, Slovakia, Alfa, 1991

Selection of method for computer-based circuits simulation.

[1] *Jozef DZIAK (1st year)*
Supervisor: [2] *Iveta TOMČÍKOVÁ*

[1,2] Dept. of Theoretical Electrical Engineering and Electrical Measurement, FEI TU of Košice, Slovak Republic

[1] jozef.dziak@tuke.sk, [2] iveta.tomcikova@tuke.sk

***Abstract*—Paper deals with several circuit analyses (Superposition theorem, Kirchhoff's law analysis, Nodal analysis, Mesh analysis, Modified nodal analysis, Sparse tableau analysis). It describes fundamental principle of each analysis and defines fundamental criteria for computer simulation of electrical circuits. Based on the criteria the options for using each method for computer simulation was explored and the appropriate method for computer simulation was selected.**

***Keywords*—circuits simulation, circuit analysis, computer-based simulation**

I. Introduction

Circuit simulation is a technique for checking and verifying the design of electrical and electronic circuits and systems prior to manufacturing and deployment. It is used across a wide spectrum of applications, ranging from integrated circuits and microelectronics to electrical power distribution networks and power electronics. Circuit simulation combines mathematical modeling of the circuit elements, or devices, formulation of the circuit equations and techniques for solution of these equations. We will focus mainly on the formulation and solution of the network equations in this paper [1].

Many different kinds of network element are encountered in network analysis. For circuit analysis it is necessary to formulate equations for circuits containing as many different types of network elements as possible. There are various methods for formulation of circuit equation. Set of circuit equations consist of connection equations (based on Kirchhoff's laws) and element equations.

Now we formulate the basic criteria for computer simulation of circuits. We will use them for exploring of the methods used for the simulation and selection of method for circuit analysis with computer.

In a computer program the equations have to be formulated automatically in a simple, comprehensive manner. Once formulated, the system of equations has to be solved. There are two main aspects to be considered when choosing algorithms for this purpose: accuracy and speed. In principle, accuracy depends on mathematical modeling of the circuit elements and speed depends on method. Paper deals with methods, therefore we exclude accuracy.

II. Superposition theorem

Superposition theorem (ST) designed to simplify networks containing two or more sources. It states that in a network containing more than one source, the current at any one point is equal to the algebraic sum of the currents produced by each source acting separately (viz. Fig. 1) [2]:

$$I_B = \sum_{i=1}^{n} I_i \qquad (1)$$

I_B – Matrix of branch currents
I_i – Matrix of currents produced by each source
n – Number of sources in circuit

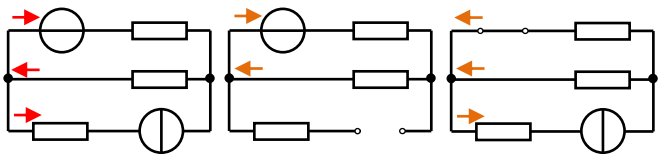


Fig. 1. Superposition theorem example

ST can only be used in linear circuits. However, the superposition theorem can be used in nonlinear resistive circuits. For any resistive network with n-independent nodes and one two-terminal nonlinear resistor, we can separate the nonlinear resistor from the network as shown in Fig. 2. We separate the nonlinear resistor R from the network, and substitute R by an independent current source and by an independent voltage source. Then, we can use the substitution theorem to change the network equivalently as long as the circuit has a unique solution [3].

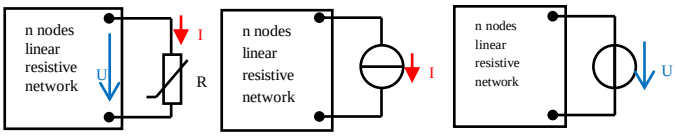


Fig. 2. Separate nonlinear resistor R, substituting R by a voltage source and substituting R by a current source

If we use ST, then circuit equations is not possible to formulate automatically. Automatic formulation of equations is a condition for computer-based circuit simulation. Therefore we will not use this method for simulation.

III. KIRCHHOFF'S LAW ANALYSIS

In Kirchhoff's law analysis (KLA) are used directly Kirchhoff's laws. Circuit equations consist of two sets of equations. Connection equations formed by using Kirchhoff's voltage law (KVL) and Kirchhoff's current law (KCL) and element equations described the V-C relationship of elements. Aim of KLA is do find all unknown's branch currents (viz. Fig. 3) [2] [4]:

$$I_{BU} = -\begin{bmatrix} A_N & \\ A_M & Z_B \end{bmatrix}^{-1} \begin{bmatrix} A_N & S_I \\ A_M & S_U \end{bmatrix} \quad (2)$$

I_{BU} – Matrix of unknowns branch currents
S_I – Matrix of currents sources
S_U – Matrix of voltages sources
Z_B – Matrix of branch impedance
A_N – Nodes incidence matrix (KCL)
A_M – Mesh incidence matrix (KVL)

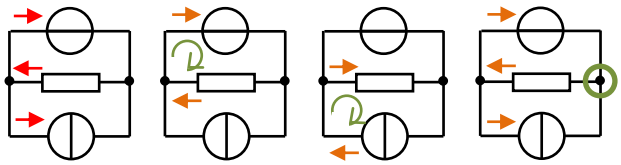


Fig. 3. Kirchhoff's law analysis example

Circuit description consists of relations between the branches (elements) and nodes corresponding A_N, and relations between branches and meshes corresponding A_M. It complicates and prolongs the computation and thus adversely affects the speed of the algorithm.

IV. NODAL ANALYSIS

The nodal analysis (NA) is based on the following idea. Instead of solving for circuit variables, current and voltage of each element, we solve for a different set of parameters, node voltages in this case, which automatically satisfy KVL. We do not need to write KVL equations and need only to solve KCL equations. The aim of nodal analysis is to determine the voltage at each node relative to the reference node (viz. Fig. 4) [2] [4]:

$$U_N = Y_N^{-1} \cdot \left(A_N S_I + A_N Y_B S_U \right) \quad (3)$$

U_N – Matrix of node voltages
Y_N – Matrix of nodes susceptance
Y_B – Matrix of branches susceptance

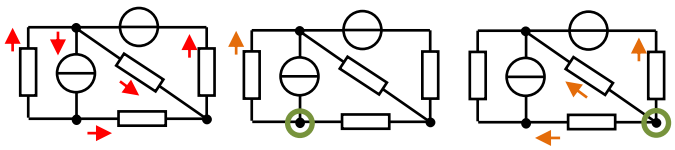


Fig. 4. Nodal analysis example

NA is a fairly comprehensive method of solving circuits. It is also suitable for computer-based simulation, but is has one complication. In NA is voltage source converted to current source. Conversion of real voltage source is possible, but ideal voltage source convert to current source is not possible.

V. MESH ANALYSIS

The mesh analysis (MA) is analogy of the NA. We solve for a new set of variables, mesh currents that automatically satisfy KCL. As such, mesh analysis reduces circuit solution to writing KVL equations. The aim of mesh analysis is to determine the currents at each independent mesh (viz. Fig. 5) [2] [4]:

$$I_M = Z_M^{-1} \left(A_M S_U + A_M Z_B S_I \right) \quad (4)$$

I_M – Matrix of mesh currents
Z_M – Matrix of mesh impedance

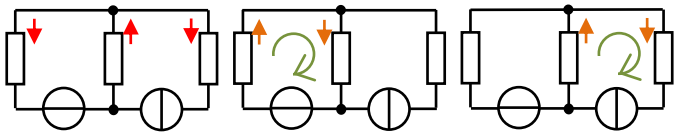


Fig. 5. Mesh analysis example

Lack of MA is that we can MA use just for planar circuits. Element or connecting wire can't intersect other element or connecting wire in plane (viz. Fig. 6). MA also has a problem with circuit description. Circuit is described by using relations between branches and meshes. Instead of simply A_N we need more complicated A_M. It adversely affects the speed of the algorithm.

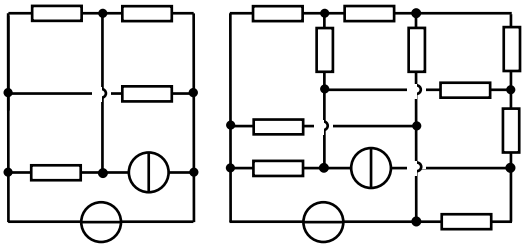


Fig. 6. Example of planar circuit and non-planar circuit

VI. MODIFIED NODAL ANALYSIS

NA is used for formulating circuit equations. However, several limitations exist in this method including the inability to process voltage sources and current dependent circuit elements in a simple and efficient manner. A modified nodal analysis (MNA) retains the simplicity and other advantages of nodal analysis while removing its limitations [5].

To handle voltage sources, the key idea is to not insist on eliminating their currents, but to retain those currents as additional variables. For of these new variables, we add a new equation, namely the branch equation for that (voltage source) element. The size of the matrix equation will grow, by as many equations as we have voltage sources, compared to nodal analysis. We can retain more currents (other currents) than just the voltage source currents [1]:

- All voltage source currents, be they independent or controlled.
- Any current that is a control variable for current control voltage source (CCVS) or current control current source (CCCS)
- Any current that is a user-specified simulation output.

We have a system of equations for the currents according NA, a system of equations for the other currents. Merging both obtained MNA system (viz. Fig. 7):

$$\begin{bmatrix} U_B \\ I_{BO} \end{bmatrix} = \begin{bmatrix} A_N Y_N A_N{}^{-1} & A_{NO} \\ -A_{NO}{}^T & Z_{NO} \end{bmatrix}^{-1} \begin{bmatrix} -A_N S_U \\ S_I \end{bmatrix} \quad (5)$$

U_B – Matrix of branch voltages
I_{BO} – Matrix of other branch currents
A_{NO} – Nodes incidence matrix of other currents
Z_{NO} – Matrix of branch impedance of other currents

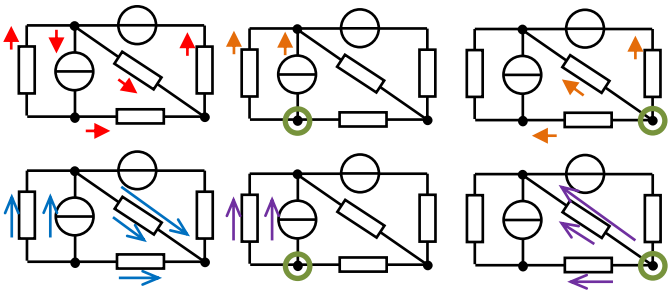


Fig. 7. Modified nodal analysis example

Level of MNA is between STA, where no currents were eliminated, and NA, where all were eliminated. This creates a disadvantage of MNA. MNA doesn't contain information about all currents and voltages in electric circuit [6].

VII. Sparse Tableau Analysis

In this approach the only matrix operation required is that of repeatedly solving linear algebraic equations of fixed sparse structure. For non-linear circuits the partial derivatives and numerical integration are done at the branch level leading to complete generality and maximum sparsity of the characteristic coefficient matrix [7].

Sparse tableau analysis (STA) described in [Hat], involves the following steps: write KCL, write KVL and write the element equations. The combination of these three sets of algebraic equations leads to the sparse tableau system [1]:

$$\begin{bmatrix} I_B \\ U_B \\ U_N \end{bmatrix} = \begin{bmatrix} A_N & 0 & 0 \\ 0 & 1 & -A_N{}^T \\ Z_B & Y_B & 0 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 0 \\ S \end{bmatrix} \quad (6)$$

I_B – Matrix of branch currents
U_N – Matrix of branch voltages
$A_N{}^T$ – Transpose nodes incidence matrix
S – Matrix of voltages sources and currents sources

This formulation has some key features consisting in fact that it can be applied to any circuit in a systematic fashion. The equations can be assembled directly from the input (circuit specification). The coefficients matrix is very sparse with mostly zero elements, although it is larger in dimension than the MNA matrix [1].

STA makes possible a simple yet truly general purpose computer program. By combining the concept of variability type with sparse matrix techniques, this program can achieve practically optimum efficiency by choosing the order of elimination so as to minimize the total operations count required for simulation and/or automated optimization [7].

VIII. Conclusion

In this paper, have been studied options of circuit analysis for computer simulation. All investigated methods of analyses could be divided into three groups. First is group of methods that are unsuitable for computer simulation, the second one is group of methods fewer suitable and the last one is group of methods that are very suitable for computer simulation.

Unsuitable for computer simulation is superposition theorem. ST represents the several analyses: Thevenin's theorem, Norton's theorem, etc. It is not possible to formulate circuit equations using these methods automatically. These methods (called "ad hoc") are inappropriate to analysis of all circuits. Instead, we need a systematic and automatic approach for formulating and solving the circuit equations.

Into group of methods fewer suitable were included Kirchhoff's laws analysis, nodal analysis and mesh analysis. They can be used for computer simulation, but their algorithm will be more complicated and / or must be used for circuits with restrictions. MA has both disadvantages. Circuit must be planar and relationships must be defined between branches and nodes and moreover between branches and meshes. KVL has the same complications as MA with circuit description. NA has problem with branches with only ideal voltage sources and current dependent circuit elements.

Unlike previous methods, the methods in the group of methods very suitable can be used for a general solution to electrical circuits. It is a modified nodal analysis and sparse tableau analysis. MNA is more compact than STA, preparation and creation of analysis is faster than STA. MNA doesn't contain information about all currents and voltages in electric circuit. Handicap of STA is that preparation of STA analysis is slower than MNA, but STA contains information about all voltages and currents.

Acknowledgment

The paper has been prepared under support of Slovak grant projects KEGA No. 005TUKE-4/2012 and KEGA 014TUKE-4/2013.

References

[1] Najm F. N. – Circuit Simulation, John Wiley & Sons, Inc., Hoboken, United States of America, 2010, ISBN 978-0-470-53871-5.

[2] Mayer D. – Úvod do teorie elektrických obvodů, SNTL / ALFA, Prague, Czech Republic, 1978, ISBN 04–536–78.

[3] Lu R., Lu A. - Applications of the Superposition Theorem to Nonlinear Resistive Circuits, APCCAS, IEEE, 2006, 1-4244-0387-1/06, 2006.

[4] Šimko V., Kováč D.: Učebné texty z teoretickej elektrotechniky I, elfa, Košice, Slovakia, 1998, ISBN 80-88786-79-7.

[5] Ho Ch., Ruehli A. E., Brennan P.A. – The modified Nodal Approach to Network Analysis. IEEE Transactions on circuit and systems, 1975

[6] Vansáč M., Vince T. – Sparse tableau analysis of electrical circuits, XIV International PhD Workshop OWD, Poland, 2012.

[7] Hachtel G.D, Brayton R.K., Gustavson F.G.: The Sparse Tableau Approach to Network Analysis and Design. IEEE Transactions on circuit theory, 1971.

Simulation and Comparison of Microstrip Bandstop Filters with L-resonators

[1]*Tibor ROVENSKÝ (1st year),* [2]*Kornel RUMAN (2nd year)*
Supervisor: [3]*Alena PIETRIKOVÁ*

[1]Dept. of Technologies in Electronics, FEI TU of Košice, Slovak Republic

[1]tibor.rovensky@tuke.sk, [2]kornel.ruman@tuke.sk, [3]alena.pietrikova@tuke.sk

***Abstract*— This paper deals with design and simulation of microstrip bandstop filters with L-resonators. In this article two types of narrowband notch filters are simulated, one filter with three L-resonators and another one with five L-resonators. For the purposes of simulation Ansoft designer from ANSYS was used. Their scattering parameters forward transmission coefficient (S21) and input reflection coefficient (S11) from 2.3 GHz to 2.6 GHz are compared. Microstrip filters were simulated using RT/duroid 6010LM microwave ceramic-PTFE composite laminate from ROGERS Corporation.**

***Keywords*— bandstop filter, half wavelength L-resonators, microstrip, Ultra-Wide-Band radar systems**

I. INTRODUCTION

UWB radar systems are used in a various applications such as auto-electronics, detection of subject behind the wall, medical applications etc. UWB are well-known for usefulness in sensor implementations as well as in localization or communication services [1]. However, big challenge is to design and construct precise filters to remove unwanted signals from environment. Nowadays, there are signals with various frequencies everywhere, especially signal in Wi-Fi area from 2.4 GHz to 2.5 GHz [2]. By filtering signals on receiver and transducer antennas UWB can achieve results with higher accuracy.

Quality of filters and design itself depends on dielectric properties of substrates. Dielectric constant and loss tangent have significant influence on microstrip structures. Dielectric substrate RT/duroid 6010LM from ROGERS Corporation with copper foil on both sides was used for simulation and design. This material was developed for the purposes of microwave circuits [3].

The aim is to investigate behavior of two types of filters with different number of L-resonators, which provides coupling between main transmission line and themselves to attenuate signals with tuned frequency. By comparison of scattering parameters (S11 and S21) of both filters we can decide which type of designed microstrips structure is better for our purposes from two points of view. The first is to meet our requirements of minimal attenuation and bandstop bandwidth and the second one is to minimalized size of fabricated devices as well as costs of devices. Result of this analysis will be used in further filter designing to fabricate low cost devices with high stability.

II. FILTER DESIGN

A. Substrate

Before designing filters the most important thing is to know properties of used dielectric material. RT/duroid 6010LM was used for simulation of designed filters. It is microwave laminate developed for electric and microwave circuit applications. Microwave laminate feature ease of fabrication and stability in use. It has tight dielectric constant and thickness control, good thermal mechanical stability and low moisture absorption. Electrodeposited copper foil is on both sides. According to the high dielectric constant it provides an option for designers to reduce size of boards. Characteristic of substrate Rogers RT/duroid 6010LM is shown in Table 1.

TABLE 1
ROGERS RT/DUROID 6010LM SUBSTRATE SPECIFICATION

Property	Symbol	ROGERS RO4003C
Dielectric Constant	ε_r	10,2
Dissipation Factor Tan	Δ	0,0023
Dielectric Height	H	0,025" (0,632 mm)
Resistivity Compared to Copper	P	1
Metal Thickness	M	½ oz. (17μm)
Volume Resistivity	-	5 x 10^6 (MΩ•cm)
Surface Resistivity	-	5 x 10^5 (MΩ•cm)
Moisture Absorption	-	0.01 (%)
Melting temperature Tg	-	>260 (°C DSC)

B. Microstrip structure

Many structures for bandstop filters exist [4], [5]. To meet requirements of narrow banstop only in the area from 2.4 GHz to 2.5 GHz we designed filters with half wavelength L-resonators. Filter with five L-resonators is shown in the Figure 1. In the Figure 2 filter with three L-resonators is presented.

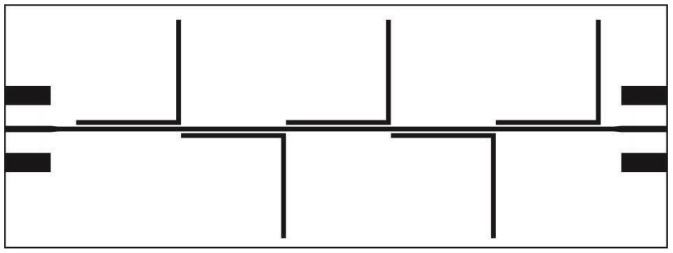


Fig. 1. Microstrip bandstop filter with five L-resonators and pads for connector.

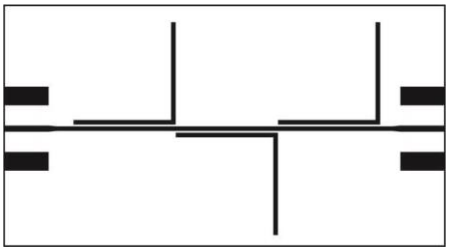


Fig. 2. Microstrip bandstop filter with three L-resonators and pads for connector.

Length of L-resonators is half wavelength of guided wave. Half wavelength was calculated using these equations [6]:

$$\lambda_0 = \frac{c}{f} \quad (1)$$

where λ_0 is free space wavelength, c is speed of light, f is used center frequency (2.45 GHz)

$$\lambda_g = \frac{\lambda_0}{\sqrt{\varepsilon}} \quad (2)$$

λ_g is guided wave wavelength and $\mathcal{E}$ is permittivity of used dielectric substrate. Length of resonators was slightly tuned and spacing of main line and resonators was fitted to meet requirements of filters characteristic. Width of microstip structure is calculated to 50 ohm characteristic impedance [7], [8].

III. Simulation and Results

A. *Ansoft Designer*

Nowadays, there are a lot of software tools for simulating microstip structures. According to various options how to composed microstips we used Ansoft Designer from ANSYS. This programme features libraries with constituent elements of microstrips such as transmission line, bends, coupled resonators etc. and libraries with many types of materials for the purposes of simulations what is very useful. It provides two types of simulations simple circuit simulation and planar electromagnetic simulation which includes far and near field radiation, currents density and many other parameters. Important feature of this software is exporting microstrip structures to .dxf or .ger files what is very useful for fabrication [9].

B. *Simulations*

For making precise simulation we set parameters of used dielectric substrate (permittivity, loss tangent, height of dielectric and conductor traces) which is needed to calculate width of microstrips by tool implemented in this software. To investigate and verify behavior of designed filters there is necessity to simulate scattering parameters and to illustrate these parameters in frequency domain characteristic.

Simulations were done from 2.3 GHz to 2.6 GHz frequency range. We examined frequency bandwidth of stopband characteristic and attenuation of this stopband. Requirements were to design filter with stopband from 2.4 GHz to 2.5 GHz (bandwidth 0.1 GHZ) and minimal attenuation -40 dB.

In the Figure 3 input reflection cofficient is shown. There are small differences between reflection coefficients of both filter types. Filter with five L-resonators has wider reflection coefficient than filter with only three resonators which can be caused by more coupled resonators.

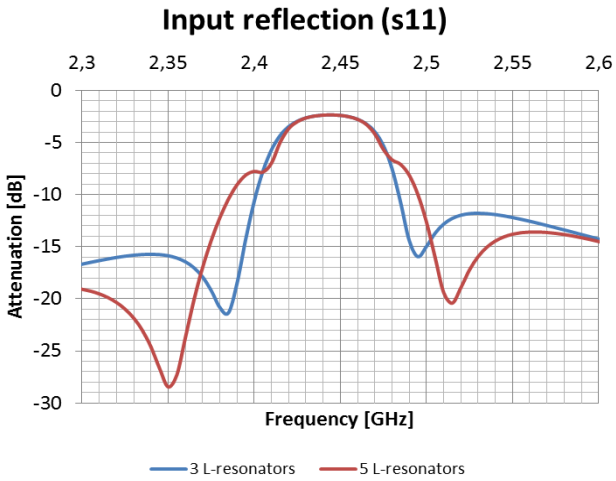


Fig. 3. Input reflection coefficient (s11) of filters with three and five L-resonators from 2.3 GHz to 2.6 GHz.

In the Figure 4 forward transmission coefficient is shown. Differences between frequency domain characteristics are significant. Filter with three resonators has attenuation in stopband -50 dB. Attenuation of filter with five resonators is almost two times higher than filter with only three couplings and it is -87 dB. Shape of both curves is absolutely smooth with no deformations caused by errors made during calculations and filter design.

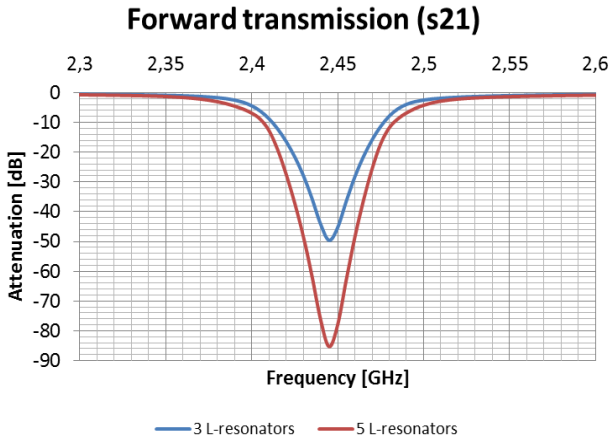


Fig. 4. Forward transmission coefficient (s11) of filters with three and five L-resonators from 2.3 GHz to 2.6 GHz.

In the Figure 5 stopband of notch filter is shown. Differences between stopband bandwidth of two filters is not very enormous but they could have major impact on UWB radar system. For our purposes of filtering signals only with frequencies from 2.4 GHz to 2.5 GHz it could caused removing not only unwanted signal which can influenced acuraccy of UWB, but removing useful information too. Filter with three resonators has stopband characteristic only in bandwidth which is wanted to be removed from signal and then processed by UWB. Microstrip filter with five resonators has bandwidth of band with bandstop characteristic from 2.38 GHz to 2.51 GHz. Removing useful part of signal can caused decreasing of UWB radar system acuraccy.

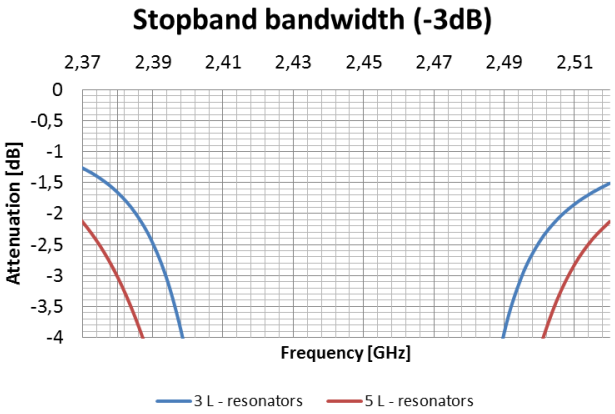


Fig. 5. Stopband bandwidth (-3dB) of filters with three and five L-resonators.

Both filter types reach required parameters of minimal attenuation and bandstop bandwidth. Filter with five L-resonators achieve attenuation twice that of requirements and bandstop bandwidth of this filter is greater than required, it could decreased accuracy of UWB. Three L-resonators filter exactly meet requirements of bandstop bandwidth and attenuation is quite higher.

IV. Conclusion

The bandstop filters with L-resonators were designed and simulated in high frequency area from 2.3 GHz to 2.6 GHz. Analyses of bandstop bandwidth and attenuation reveals that filter with three L-resonators meet our requirements and even size of this device is smaller which reduce costs of fabrication. Designed microstrips structure will be fabricated, measured by VNA (Vector Network Analyzer) and results will be included in future work.

Acknowledgment

This work was supported by VEGA Grant No. 1/0218/13.

References

[1] I. J. Immoreev, and J. D. Taylor, "Future of radars," Ultra Wideband Systems and Technologies, 2002. pp. 197-199.
[2] Wi-Fi / WLAN Channels, Frequencies and Bandwidths, Available on the Internet: 8.3.2013, http://www.radio-electronics.com/info/wireless/wi-fi/80211-channels-number-frequencies-bandwidth.php
[3] RT/duroid 6006/6010LM High Frequency Laminates, Available on the Internet: 8.3.2013, http://www.rogerscorp.com/documents/612/acm/RT-duroid-6006-6010-laminate-data-sheet
[4] Jia-Sheng Hong, "Microstip Filters for RF/Microwave Aplications," John Wiley & Sons, Inc., Hoboken, 2nd edition, New Jersey, 2011, 655.
[5] Jia-Sheng Hong and M. J. Lancaster, "Microstip Filters for RF/Microwave Aplications," John Wiley & Sons, Inc., Hoboken, New York, ISBN 0-471-22161-9, 2001, 471.
[6] Joseph F. White, "High frequency techniques: An Introduction to RF and Microwave Engineering," John Wiley & Sons, Inc., 2004, 528.
[7] There's Nothing Magic About 50 Ohms, Available on the Internet: 8.3.2013, http://www.highfrequencyelectronics.com/Archives/Jun07/HFE0607_Editorial.pdf
[8] Width and Effective Dielectric Constant Equations for Design of Microstrip Transmission Lines, Available on the Internet: 8.3.2013, http://www.rogerscorp.com/documents/780/acm/Design-Data-for-Microstrip-Transmission-Lines-on-RT-duroid-Laminates.pdf
[9] ANSYS DesignerRF, Available on the Internet: 8.3.2013, http://www.ansoft.com/products/hf/ansoft_designer

Solid State ^{13}C NMR Study of Changes in Physical Properties of Polypropylene Films

[1]*Magdaléna UHRÍNOVÁ (3rd year),* [2]*Peter DURANKA (3rd year)*
Supervisor: [3]*Dušan OLČÁK*

[1,2,3]Dept. of Physics, FEI TU of Košice, Slovak Republic

[1]magdalena.uhrinova@tuke.sk, [2]peter.duranka@tuke.sk, [3]dusan.olcak@tuke.sk

***Abstract* — The physical aging of biaxially oriented polypropylene films was studied by the measurements of high-resolution solid state ^{13}C NMR spectra for the samples which were aged at the stable temperature of 23 °C for 15,150, 390, 620 and 660 days. The spectra were measured at 30 and 98 °C. The presence of the α-crystalline form in the polypropylene film was identified by the NMR spectra at the beginning of aging. From the spectra recorded for aged samples can be deduced that physical aging includes the rearrangements of the crystalline chains and transformation of the α-form into hexagonal one. The data obtained from the deconvolution of the spectra measured at 98 °C indicate that physical aging does not influence the distribution of the chain conformations in amorphous regions. On the other hand, the chain rearrangements and broadening of the chain conformations distribution within crystalline regions due to physical aging could be drawn from the data. The primary crystallinity of as-supplied sample was 0.58, after 620 days of aging the crystallinity was observed to increase to the value of 0.63.**

***Keywords* — physical aging, isotactic polypropylene, ^{13}C solid-state NMR**

I. Introduction

Chemically diverse range of materials that includes polymers, inorganic glasses, composites and amorphous metals usually exists in a nonequilibrium state. There is a slow and gradual approach to equilibrium in these materials, which changes their properties. The relaxation process from the nonequilibrium state toward equilibrium is commonly referred to as physical aging. Physical aging, which relates to structural relaxation of the glassy state, is accompanied by changes in almost all physical properties, and is considered to be a basic feature of the glassy state. Glassy state relaxation process shows memory effect, that is, relaxation from a particular state depends not only what that state is, but also on how that state was reached. While physical aging affects the properties of many different types of glasses, these effects have been most thoroughly investigated for polymer glasses because of their practical importance. The rate of physical aging is determined in part by how far below the glass transition temperature the material is used and it is known that the glass transition temperature values are lower for polymers than for most inorganic glasses. [1-3].

Physical aging of partially crystalline polymers refers to the evolution toward an equilibrium state and associated changes in the morphological organization and physical properties at temperatures between the glass transition temperature and melting temperature. It depends on temperature and, generally, it involves the densification of amorphous fraction, restrictions on the chain mobility, crystallization of constrained polymer chains, lamellar thickening and other processes. Amorphous regions of partially crystalline polymers have the potential to crystallize and to increase crystallinity of the polymer during aging. In addition to the lamellae formed during primary crystallization also new lamellae can be formed during secondary crystallization. The thickness of the newly-formed lamellae depends on the aging temperature with respect to the glass transition temperature and also on the initial morphology of the partially crystalline polymer [4].

The aim of this paper is to study the structural rearrangements that accompany physical aging of biaxially stretched polypropylene films using MAS ^{13}C NMR spectroscopy. The influence of aging time on the phase composition, crystallinity and crystalline structure of the samples is also studied.

II. Experimental

The high-resolution ^{13}C solid state NMR measurements were carried out on the biaxially stretched polypropylene films produced by Chemosvit a.s., Slovakia. The ^{13}C MAS NMR spectra were measured for the samples stored in the NMR laboratory at the stable temperature of 23 °C. Strips of the 12 mm width were cut from the polypropylene films and then they were rolled into the cylindrical form and inserted into the rotor.

The NMR experiments on the samples aged for 15 and 660 days were performed at ambient probe temperature, which was of 30 °C. The NMR experiments for the samples aged for 15, 150, 390 and 620 days were performed at 98 °C. The measured samples are denoted as iPP-t_a where t_a means the aging time in days, i.e. the time which elapsed up to the starting of NMR experiments that took approximately one weak.

The ^{13}C MAS NMR spectra were measured on the solid state Varian NMR spectrometer equipped with an actively shielded wide-bore magnet generating magnetic field of 9.4 T. The resonant frequency of the ^{13}C nuclei corresponding to the mentioned magnetic field is about 100 MHz. Experiments were carried out with a probe-head using the 4 mm rotor under spinning at the rate of 10 kHz. The duration of the ^{13}C excitation pulse, high power proton decoupling and recycle delay were 2.7 μs, 100 kHz and 240 s, respectively. Deconvolutions of the measured high-resolution NMR spectra were done using the Lorentzian lineshapes and the line intensities and chemical shifts were obtained from the deconvoluted spectra. The chemical shifts were referenced to the TMS using adamantane as an external standard.

III. Results

Fig. 1 shows the ^{13}C NMR spectra measured at 30 °C for the iPP-15 and iPP-660 samples. The common feature of both spectra is the existence of three peaks with the chemical shifts of about 22.0, 26.5 and 44.3 ppm, which can be assigned to the carbons in methyl CH_3, methine CH and methylene CH_2 groups, respectively [5]. The differences between the spectra can be observed in the shape and width of the peaks related to the particular groups. The spectrum for iPP-15 exhibits one narrow peak related to the methine groups and two doublets associated with methylene and methyl groups. Evidently, physical aging results in disappearing of the doublets and broadening of all three peaks.

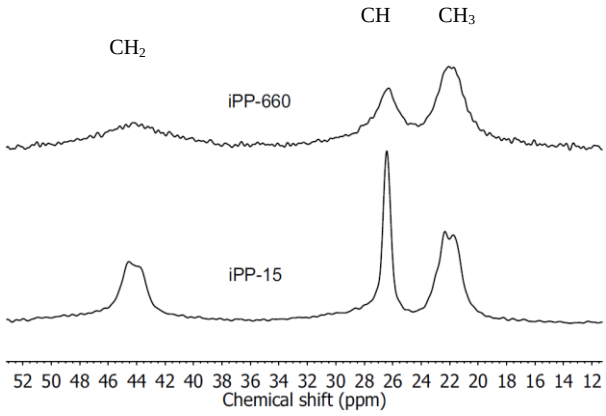


Fig. 1 The high-resolution solid state ^{13}C NMR spectra measured for the samples iPP-15 and iPP-660 at 30 °C

The ^{13}C NMR spectra for iPP-15 measured at 30 and 98 °C are shown in Fig. 2. It can be seen that besides the lines observed at 30 °C, the new resonance lines of CH_2 and CH carbons appear in the spectrum measured at 98 °C. Similar effect was also observed in the spectra for iPP prepared in pelletized form where these new lines were assigned to the carbons within the chains of amorphous regions [5, 6]. We assume that the same is true in the case of iPP films and the peaks associated with carbons in crystalline as well as amorphous regions are resolved in the spectrum measured at 98 °C. The shape of the CH_3 peak was changed by the temperature increase. As was mentioned above this peak was split into two lines with about the same amplitudes at 30 °C. Probably due to the change of the line widths of these lines at 98 °C the peak with a shoulder is observed at this temperature.

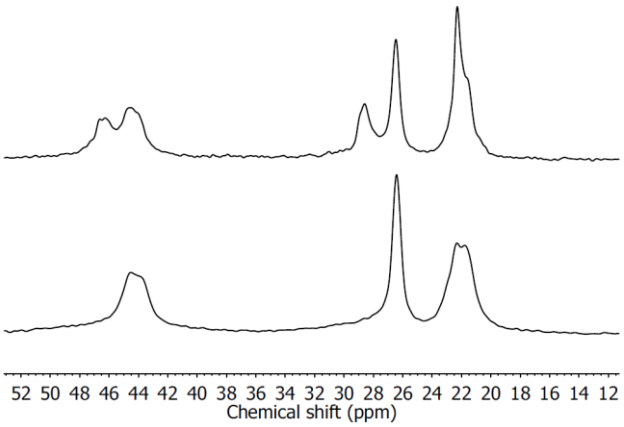


Fig. 2 The high-resolution solid state ^{13}C NMR spectra measured for the sample iPP-15 at 30 (down) and 98 °C (up)

The effect of aging process on the amorphous and crystalline regions of polypropylene films can be observed in the spectra measured for the samples after different aging times t_a (Fig. 3). The behavior of CH_2 and CH peaks is very similar with increasing aging time t_a. The shapes and amplitudes of the amorphous CH_2 and CH peaks are not influenced by the aging process. On the other hand, physical aging dramatically affects crystalline regions of the polypropylene films which can be deduced from both the decrease of amplitudes and broadening of the crystalline CH_2 and CH peaks with increasing aging time. Besides, the small change of the shape of the CH_3 peaks can be also deduced from the spectra in Fig. 3. Aging process has tendency to smooth the shoulder in the CH_3 line.

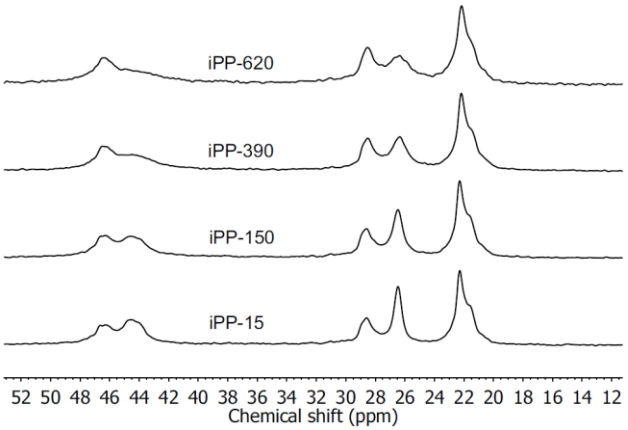


Fig. 3 The high-resolution solid state ^{13}C NMR spectra measured for the samples as indicated at 98 °C

IV. Discussion

Isotactic polypropylenes are highly crystalline polymers with crystalline domains embedded into amorphous matrix. The crystalline regions are known to adopt a helical molecular chain conformation in the solid state with three monomer units per turn. The α-crystalline form consists of iPP chains

packed in monoclinic unit cell and the most likely structure contains left and right handed pairs of chains in close proximity. On the basis of the arrangement of the molecular chains in a unit cell the existence of two different packing sites in crystals with two distinct environments for the monomer units are expected in this structure. The hexagonal β-forms of iPP contains helical chains, which are arranged in groups of the same helical handedness resulting in the distant packing of left- and right-handed chains. The smectic form of iPP is only partially ordered compared to the α- and β-forms, though the iPP chains remain in the same helical conformation. Smectic iPP is considered to be primarily disordered in the intermolecular packing of its chains [7].

The already published high-resolution solid-state ^{13}C NMR studies of isotactic polypropylene polymorphs showed that the shape of the ^{13}C NMR spectra measured at room temperature depends on the chain arangement in the crystalline lattice. The methyl and methylene peaks in α-crystalline form were observed to split by about 1 ppm and doublets with intensities of 2:1 appeared in both peaks. The doublets were suggested to reflect the presence of two different packing sites in α-crystalline form. On the other hand, no doublets were observed in the spectra of the β-crystalline and smectic forms and carbons of each group have been shown to exhibit a single symmetrical resonance. It was inferred from the spectra that disappearance of the methyl and methylene doublets was due to the absence of different packing sites [7, 8].

TABLE I
WIDTHS AND INTENSITIES OF THE METHINE PEAKS RELATED TO THE AMORPHOUS AND CRYSTALLINE REGIONS OF THE SAMPLES OBTAINED FROM THE DECONVOLUTIONS OF MAS ^{13}C NMR SPECTRA MEASURED AT 98 °C

	Amorphous		Crystalline	
	Width (Hz)	I (%)	Width (Hz)	I (%)
iPP-15	105	42	64	58
iPP-150	100	42	81	58
iPP-390	100	44	122	56
iPP-620	100	37	220	63

As was mentioned above, the doublets in methyl and methylene peaks in the high-resolution solid-state ^{13}C NMR spectrum for iPP-15 measured at 30 °C (Fig. 1) result from the presence of the α-crystalline form in the studied biaxially oriented polypropylene film. However, the ratio of the line intensities in the doublets, which is obviously not equal to 2:1, as would be expected if only the α-crystalline form was present (the ratio 1.35:1 follows from the deconvolution), indicates that except of α-form other crystalline modification (hexagonal, smectic) can exist in the sample. The disappearance of the doublets in the spectrum measured for iPP-660 at 30 °C can be explained by the rearrangements of the crystalline chains and transformation of the α-form into hexagonal one during physical aging. Structureless and very broad methylene, methine and methyl peaks in the spectrum measured for iPP-660 reflect the broad chain conformations distribution in crystalline and amorphous regions. The blend of the diverse crystalline modifications can be one of the reasons of the line-broadening.

The measurements of the high-resolution solid-state ^{13}C NMR spectra at elevated temperatures make it possible to study the effect of physical aging in the amorphous and crystalline regions. Owing to the fact that the best resolution of the amorphous and crystalline peaks was reached for the methine groups the deconvolution of this part of the spectra was performed and the results are summarized in the Tab. 1. From not changing line widths of this peak can be inferred that physical aging does not influence the chain conformations in amorphous regions. On the other hand, strong dependence of the linewidth of the peak produced by carbons in crystalline region on the aging time supports our conclusion concerning chain rearrangements within crystalline regions and the broadening of the chain conformations distribution by the physical aging process.

Crystallinities of the biaxially oriented polypropylene films were determined from the intensities of the crystalline methine lines obtained from deconvolution (Tab.1). It can be seen that primary crystallinity of the sample was 0.58, the aging for 620 days resulted in the increase of crystallinity to the value 0.63.

REFERENCES

[1] L. C. E. Struik, "Physical Aging in Plastics and Other Glassy Materials," in *Polymer Engineering and Science*, Vol. 17, 1977, pp. 165-173.
[2] I. M. Hodge, "Physical Aging in Polymer Glasses," in *Science*, Vol. 267, 1995, pp. 1945-1947.
[3] H. N. Lee, M. D. Ediger, "Interaction between physical aging, deformation, and segmental mobility in poly(methyl methacrylate) glasses," in *The Journal of Chemical Physics*, Vol. 133, 2010
[4] C. Hedesiu, D. E. Demco, *Macromolecular Materials and Engineering*, Vol. 293, 2008, pp. 847-857.
[5] O. Fričová, "High-resolution solid-state NMR study of isotactic polypropylenes," in *Express Polymer Letters*, Vol. 6, 2012, pp. 204-212.
[6] R. Kitamaru, "Phase Structure of Polyethylene and Other Crystalline Polymers by Solid-State ^{13}C NMR," in *Advances in Polymer Science*, Vol. 137, 1998, pp. 41-101.
[7] M. A. Gomez, "High-Resolution SolidState^{13}C Nuclear Magnetic Resonance Study of Isotactic Polypropylene Polymorphs," in *Polymer*, Vol. 28, 1987, pp. 2227-2232.
[8] A. Bunn, "High-Resolution ^{13}C NMR Spectra of Solid Isotactic Polypropylene," in *Polymer*, Vol. 23, 1982, pp. 694-698.

Static alignment of the traces measured on microcontroller with accelerated 8051 core

[1]*Martin PETRVALSKY (1st year), Supervisor:* [2]*Milos DRUTAROVSKY*

[1,2]Dept. of Electronics and Multimedia Communications, FEI TU of Kosice, Slovak Republic

[1]martin.petrvalsky@tuke.sk, [2]milos.drutarovsky@tuke.sk

***Abstract*—This paper presents various algorithms focused on static alignment of power consumption traces measured on 8051-based microcontroller with accelerated core. Alignment of the traces forms important part of Differential Power Analysis (DPA) attack. The paper discusses and compares several approaches of the power consumption traces alignment suited for the microcontroller with accelerated 8051 core. This work shows the results of DPA attack using different static alignment algorithms and it also compares efficiency of the algorithms when they are applied on traces with increased level of noise. The main result of the paper is universal algorithm which can align even traces with high level of noise measured on the microcontroller with 8051-based accelerated core.**

***Keywords*—accelerated 8051, alignment of traces, DPA attack, static alignment**

I. Introduction

DPA attack [1] is the most common side-channel attack. It requires multiple power consumption measurements while the attacked device is enciphering known input plaintexts. Measured traces are subsequently used for correlation analysis which needs to have power consumption samples of corresponding data operations aligned. However the traces are often misaligned which is caused by jitter of the used oscillator or it can be done intentionally as a DPA countermeasure [2], [3].

Methods of trace alignment [4], [5], [6], [7], [8] can be divided into two categories. The first method (proposed in [4]) is called static alignment. It finds a reference in measured power consumption samples and it aligns the traces using these references. The second approach is elastic alignemnt [8]. It uses Dynamic Time Warping (DTW) and one reference trace.

In our paper we discuss only static alignment algorithms. They are faster, easier to implement and sufficient for our implementation of DPA attack. We use 8051-based microcontroller ADuC842 [9] with straightforward implementation of Advanced Encryption Standard (AES) [10] as the attacked cryptographic device.

The paper is organized as follows. In the second chapter, the DPA attack methodology is discussed. The third section presents specifications of used processors. Fourth chapter is focused on the results of the traces alignment algorithms and in the final section reader can find conclusion.

II. DPA attack

Basic principle of DPA which is depicted in Fig. 1 works as follows. Firstly, attackers input set of known data to device under test (DUT) and measure instantaneous power consumption while the processor is ciphering the data. Secondly, they create hypothesis of power consumption for all possible keys or key parts (usually bytes) using the same input data. Finally, the hypothesis which has the highest correlation with preprocessed measured traces leads to secret key.

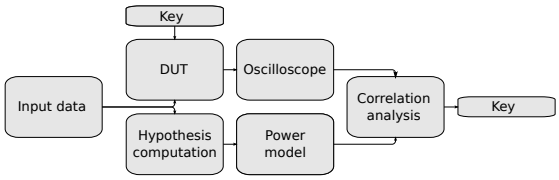


Fig. 1. Workflow of DPA attack. It has two branches. The upper one represents measurement of the power consumption. The lower branch illustrates creation of the consumption hypothesis. The final block compares measured traces with hypothesis. Hypothesis with the highest correlation leads us to secret key.

We perform DPA attack on the microcontroller which si fitted on custom evaluation board created in our department. First of all we prepare the measurement setup in order to measure instantaneous power consumption of the attacked device. Measurements are performed by digital oscilloscope Agilent Technologies DSO9404A [11] which is capable to send input data to DUT, measure its instantaneous power consumption and store measured power traces at the same time. Setup is designed to suppress noise and avoid any interference (e.g. BNC connectors, shielded cables). These factors can significantly affect results of the measurement or make the attack impossible.

In next step we load implementation of AES and secret keys to flash memory of ADuC microcontroller which are fitted on custom evaluation boards. These boards have to be modified for measurement purposes. There are various measure points on the boards which can provide different results. Traces aquisition is depicted in Fig. 2.

After measurement process, the power traces have to be preprocessed by applying data alignment. Data alignment can significantly reduce number of measured traces and improve

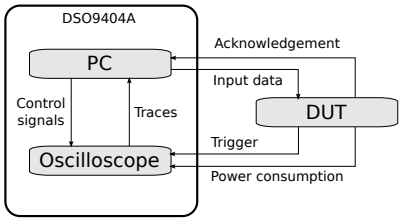


Fig. 2. Workflow of the workplace. Oscilloscope Agilent Technologies DSO9404A is used for measurements. It also contains PC which runs software that controls the whole process.

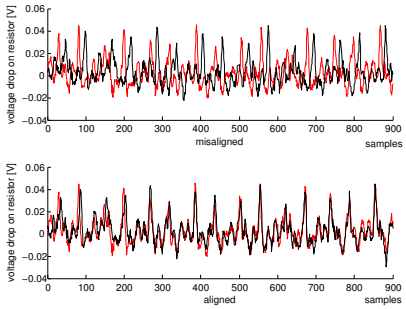


Fig. 3. Upper figure shows two misaligned traces on 900 samples. We can see that peaks do not overlap. Lower figure shows two traces after the alignment algorithm have been applied

the results. Resampling and decimation are used if needed. With statistical methods we are able to extract confidential information from data. For this purpose we use the correlation analysis between computed and measured data (1)

$$r_{H,T} = \frac{\sum_{i=1}^{N}((T_i - \bar{T})(H_i - \bar{H}))}{\sqrt{\sum_{i=1}^{N}(T_i - \bar{T})^2 \sum_{i=1}^{N}(H_i - \bar{H})^2}} \quad (1)$$

where H is power consumption hypothesis, T is vector of matching samples from every traces and N is number of traces.

III. Processor specification

There are plenty of manufacturers producing microcontrollers with accelerated 8051 core [12], [13]. The chosen target for our DPA attack is microcontroller ADuC842 [9]. It uses a 32 kHz crystal with an on-chip PLL generating core clock up to 16.78 MHz. It contains 8052-based high performance single cycle 8-bit core and various peripherals:

- 62 kB program and 4 kB data Flash/EE memory
- 256 B RAM and 2 kB extended RAM (XRAM)
- High speed 12-bit ADC and two 12-bit DAC
- DMA controller, on-chip temperature monitor
- Time interval counter (TIM), USART, I^{2}C, SPI
- Watchdog timer (WDT), power supply monitor (PSM)

The microcontroller is fitted on the custom evaluation board. This board is equipped with serial port connector for sending and receiving data. It also has prepared measure points with BNC and trigger connectors for oscilloscope probes in order to measure power consumption traces.

IV. Alignment algorithms and results

The misalignment (Fig. 3) in our case is caused by jitter of the microcontroller clock. We design 3 alignment algorithms in order to align the traces and process the aligned data to correlation analysis. All three algorithms use static alignment algorithms.

We use dataset of 256 traces each with 25,000 samples. Sampling rate is 2 GSamples/s. Measurement was done on two measurement points as shown in Fig. 4. Results of the alignment are represented by value of the correlation peak acquired from the correlation analysis.

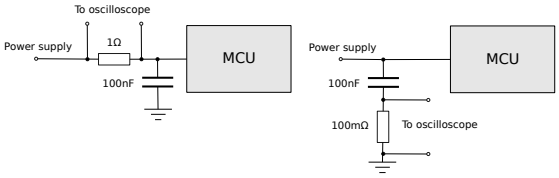


Fig. 4. Two variations of measurement point. Left circuit uses common way to measure power consumption (measurement point 1). Right circuit shows us second option of measuring on capacitor branch (measurement point 2) which can be done without differential probes. The second point gives us traces with high amount of noise which can affect the alignment algorithms.

A. Algorithm 1

The first algorithm uses rising edges of the clock peaks in traces (Fig. 3) as a reference points. We search for the rising edge with defined parameters - steepness of the edge, peak value. If we find the defined rising edge we copy several samples to new aligned trace. The number of copied samples depends on clock period. We continue this process through all the samples of all traces.

Result of this algorithm is depicted on Fig. 5. Advantages of the algorithm are simplicity and speed. Considerable disadvantage is that we need to specify the rising edge and also all rising edges in traces have to meet the specifications. If not, algorithm fails to align the traces properly and correlation peak decreases its value.

B. Algorithm 2

The second algorithm uses the clock peaks in traces as a reference points. We use differences between two samples in order to find peak. We set the first peak manually in order to synchronize the algorithm. Then we search for the next peak among samples that are 1 period away from the last peak in a search window. In our case search window is 40 samples wide. The peaks are found as zero crossing in trace with differences. The detected peak are chosen as the highest peak from found peaks. The algorithm resamples the vector of samples obtained from peak to peak to constant value (in our case 100 samples per period) and stores it in aligned trace.

Result of this algorithm is depicted on Fig. 6. The algorithm is more complicated than the first one and thus its speed is lower. On the other hand it erase disadvantages of the first algorithm and it is more effective. However the algorithm is sensitive to high amount of noise - it can hide peaks or

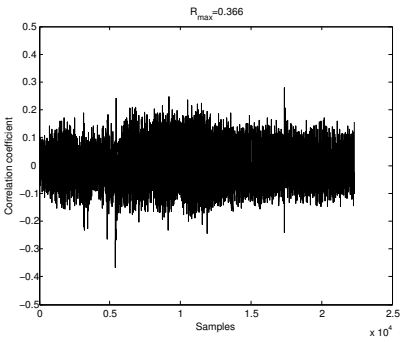


Fig. 5. Result of the first algorithm. It shows the correlation peak with absolute value 0.366 around 5,500th sample. We can see ghost peak in the second half of the trace. It is caused by random line-up of the samples in traces which partially matches the hypothesis. We use measurement point 1.

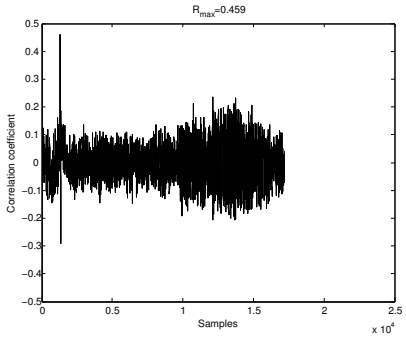


Fig. 6. Result of the second algorithm. We can see that the peak is more significant than the peak from the first algorithm. Its absolute value is 0.459 around sample 2,000. There are less samples in the figure. It is caused by choosing first peak and by resample process where each period (120 samples) was downsampled to 100 samples. We use measurement point 1.

slightly move them which can lead to local misalignment and to deterioration of the results of DPA attack.

C. Algorithm 3

The third algorithm (Fig. 7 and 8) have been developed because of need to align traces measured also on measurement point 2. These traces show high amount of noise so the first two algorithms can not be used. They do not give us any results at all when used to the traces with noise.

In the third algorithm we firstly use band pass filter to filter everything but clock frequency so we will have a trace only with clock frequency including its jitter. Then we find peaks in this filtered trace and perform alignment using the peaks. This approach solves the problem with traces with high amount of noise.

This algorithm can be used universally and it gives us the best results. It is robust against noisy traces and the only input parameter is clock frequency but on the other hand it is the most complicated algorithm and the slowest

V. Conclusion

We have developed and tested 3 different static alignment methods. The first algorithm gives us the worst results but it is easy to implement and very fast. Second algorithm significantly improves the result but the implementation is more complicated. The most complex algorithm is the third algorithm which can be deployed on traces with high amount of noise and gives us the best results.

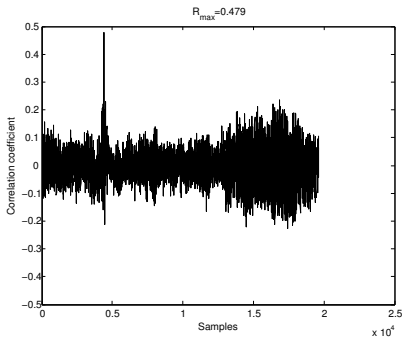


Fig. 7. Result of the third algorithm. We can see that the peak has even slightly higher value than the peak from the second algorithm. Its absolute value is 0.479 around sample 4,500. We use measurement point 1 in order to compare measurement methods.

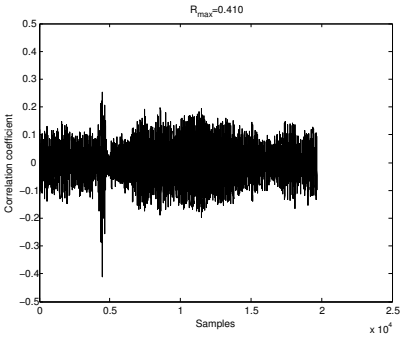


Fig. 8. The result of DPA attack on the traces measured on measurement point 2 and aligned by the third algorithm. It shows significant peak (0.410 absolute value) around sample 4,500. When we try to use first two algorithms we do not obtain any peak.

The future work can be development of the algorithms based on elastic alignment which in our opinion can perfectly align the traces and improves the results of DPA attack. The DPA countermeasures based on random insertion of delays can be taken into to the account in future alignment methods.

Acknowledgment

This research was supported by APVV-0586-11 grant. The author would like to thank assoc. prof. Milos Drutarovsky for his guidance and useful advice.

References

[1] P. C. Kocher, J. Jaffe, B. Jun, and P. Rohatgi, "Introduction to differential power analysis," *J. Cryptographic Engineering*, vol. 1, no. 1, pp. 5–27, 2011.

[2] Y. Lu, M. O'Neill, and J. Mccanny, "FPGA implementation and analysis of random delay insertion countermeasure against DPA," in *ICECE Technology, 2008. FPT 2008. International Conference on*, 2008, pp. 201–208.

[3] J.-S. Coron and I. Kizhvatov, "Analysis and improvement of the random delay countermeasure of CHES 2009," in *Cryptographic Hardware and Embedded Systems, CHES 2010, 12th International Workshop*, ser. Lecture Notes in Computer Science, vol. 6225. Springer, 2010, pp. 95–109.

[4] S. Mangard, E. Oswald, and T. Popp, *Power Analysis Attacks: Revealing the Secrets of Smart Cards (Advances in Information Security)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2007.

[5] J. Hogenboom, "Principal Component Analysis and Side-Channel Attacks," Diploma Thesis, Radboud University Nijmegen, The Netherlands, 2010.

[6] Q. Tian and S. Huss, "A general approach to power trace alignment for the assessment of side-channel resistance of hardened cryptosystems," in *Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2012 Eighth International Conference on*, July, pp. 465–470.

[7] R. Muijrers, J. Woudenberg, and L. Batina, "Ram: Rapid alignment method," in *Smart Card Research and Advanced Applications*, ser. Lecture Notes in Computer Science, E. Prouff, Ed. Springer Berlin Heidelberg, 2011, vol. 7079, pp. 266–282. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-27257-8_17

[8] J. Woudenberg, M. Witteman, and B. Bakker, "Improving differential power analysis by elastic alignment," in *Topics in Cryptology CT-RSA 2011*, ser. Lecture Notes in Computer Science, A. Kiayias, Ed. Springer Berlin Heidelberg, 2011, vol. 6558, pp. 104–119. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-19074-2_8

[9] Analog Devices, "ADuC842 datasheet and product info," http://www.analog.com/en/processors-dsp/analog-microcontrollers/aduc842/products/product.html, Accessed: 10/03/2013.

[10] J. Daemen and V. Rijmen, *The Design of Rijndael: AES - The Advanced Encryption Standard*. Berlin, Heidelberg, New York: Springer Verlag, 2002.

[11] Agilent Technologies, "DSO9404A datasheet and product info," http://www.home.agilent.com/en/pd-1632456-pn-DSO9404A/oscilloscope-4-ghz-4-analog-channels, Accessed: 10/03/2013.

[12] Intel, *MCS51 microcontroller family users manual*. [Online]. Available: http://www.industrologic.com/MCS51FamilyUsersGuide.pdf

[13] EFTON, "8051 comparison," http://www.efton.sk/51comp/51comp.htm, Accessed: 10/03/2013.

Substation protection scheme based on IEC 61850

[1]*Marian HALAJ (5st year)*, [2]*Peter HERETIK (2st year)*
Supervisor: [3]*Ladislav VARGA*, [4]*Justin MURIN*

[1,3]Dept. of Electrical Power Engineering, FEI TU of Košice, Slovak Republic
[2,4]Institute of Power and Applied Electrical Engineering, FEI STU of Bratislava, Slovak Republic

[1]marian.halaj@centrum.sk, [2]peterheretik@gmail.com, [3]ladislav.varga@tuke.sk, [4]justin.murin@stuba.sk

***Abstract*—This article deals with the main features of substation communication protocol IEC 61850 and its application in power system protection. The analyze of busbar protection model gives engineers a recommendations for the optimization of the reliability, fault clearing times, as well as hard-wiring connections.**

***Keywords*—Protection, communication, IEC 61850, GOOSE, relay, circuit breaker**

I. Introduction

Requirement for permanent monitoring, control and fault detection are some of the main features of almost all power distribution systems and of all voltage levels. Along with the keeping of power quality within the specified range system must be continuously able to detect any fault occurred and clear this one with the minimal impact on the remaining healthy system. This is so called selectivity of tripping. For a years different manufacturers have developed a lot of control and protective systems that haven't communicate or operate mutually. IEC standard 61850 allows customers to build new systems with the devices of different manufacturers as well as other advantages described further.

II. IEC 61850 Architecture

IEC 61850 was created as a internationally standardized protocol for communication and integration systems built from multivendor intelligent electronic devices (IEDs) that are interconnected to perform protection, monitoring, automation, metering, and control. This protocol enables wild range of information exchange between different IEDs, like measurements, status indication, interlocking, closing and opening orders, etc.

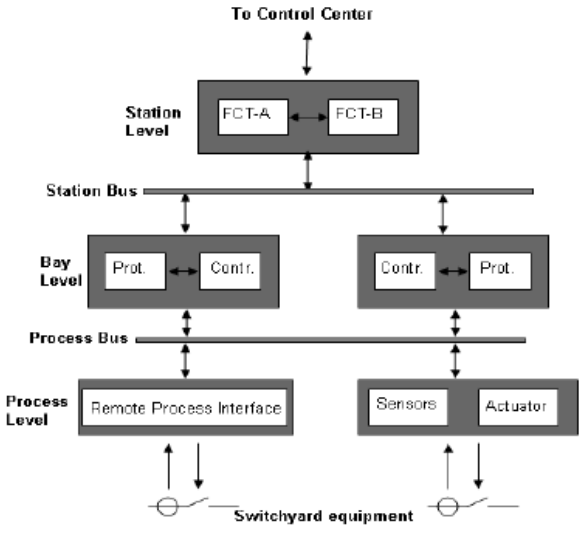


FIGURE I
IEC 61850 SUBSTATION ARCHITECTURE

From the functional point of view the IEC 61850 operates on three (FIGURE I) different levels [1]:

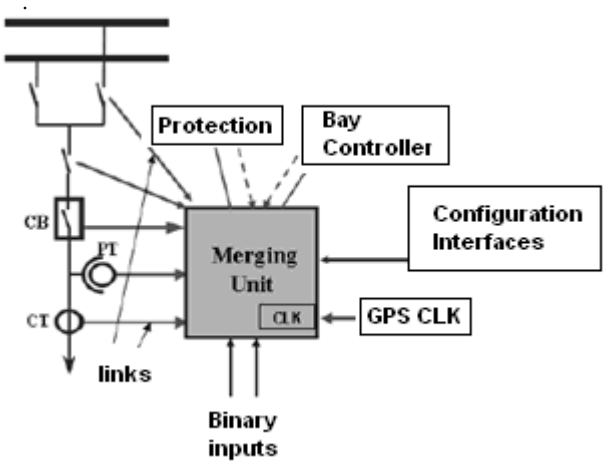


FIGURE II
PROCESS BUS CONCEPT

Process level: This level includes switchyard equipments such as CTs / PTs, Remote I/O, actuators, etc.

Bay level: Bay level includes protection and control IEDs of different bays.

Station level: The functions requiring data from more than one bay are implemented at this level.

Process bus: This facilitates the time critical communication between protection and control IED to the process (the primary equipment in the substation), such as sampled values, binary status signals or binary control signals. This layer of the substation is related to the gathering information, such as voltage, current, status indication, etc. Data transfer within this layer is available in two ways. Unidirectional multi-drop point-to-point fixed link carry a fixed dataset of the length of 16 bites. This is appropriate for transfer of sampled values. Sampled measured values (SMV) data transfer is able to transfer the values of different sizes as well as process that together.

Process bus is based on the operation of merging (FIGURE II) unit that collects primary data, time measurements and process these data to the form appropriate for a transfer, means sampling. Data are transferred to the IEDs of bay level via high-speed Ethernet network.

Station bus: It facilitates communication between station level and bay level. Station bus provides primary

communication between different logical notes. This communication includes also protection functions, control and supervision. Communication is available on connection oriented bases or GOOSE.

A. IEC 61850 Device model

According to [2] this model begins with a physical device that is typically defined by its network address (gateway). Within each physical device, there may be one or more logical devices that contains of one or more logical nodes (FIGURE III). A logical node is a group of data representing and describing a part of system, e.g. metering. This group contains all required information logically related to that group, in case of metering all measured values.

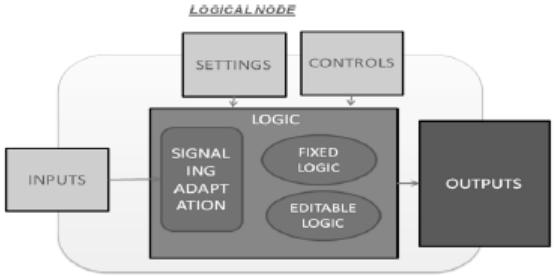


FIGURE III
STRUCTURE OF LOGICAL NODE

All levels of IEC 61850, including logical nodes, are described by substation configuration language that is based on XML.

B. Generic Object Oriented Substation Event (GOOSE)

GOOSE communication is one of the main features of IEC 61850. It enables data exchange on station bus based on the publisher-subscriber concept [3]. Comparing to another station bus protocol connection oriented base the GOOSE message may be received by several receivers. This feature was already available in older communication protocols using client-server approach as well, but in that case the communication worked on all 7 layers that is significantly time consuming. GOOSE message principle doesn't confirm the delivery receipt, but repeats the message several times instead. This feature provides very fast communication. Messages can be categorized according to importance by means of SCL [4]. Most critical messages, like tripping or blocking signals may be received within 3ms that is in certain cases faster than hard-wired connection. Fast acting functionality of these messages enables us to model the clearing of most critical faults on electrical power distribution, like short circuits on busbar.

III. PROTECTION SYSTEM MODELLING

Faults on busbars (FIGURE IV) are considered as one of the most critical defects on the power distribution systems [5]. Very high short circuit currents sometimes of hundreds of kilo-amps may cause very expensive damages on the equipment, sometimes even the risk for personnel. Such faults must be cleared within few milliseconds.

Depending on the type of protective relays used we can create three models of busbar protection:

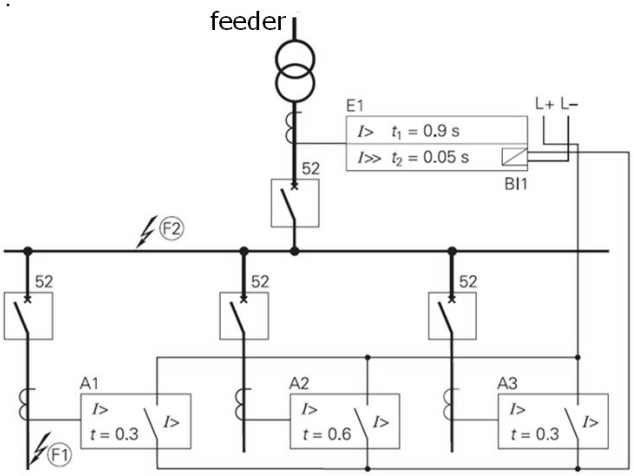


FIGURE IV
BUSBAR PROTECTION LAYOUT

- Differential protection (87B)
- Overcurrent protection (51)
- Breaker failure (50BF)

Independent of the model used the most important action in order to make sure the proper operation of the power system protection is to identify the zone where the fault occurred and keep the selectivity of tripping. This condition is important from the operation point of view, because in same cases the lost of service may cause much higher losses than the fault itself. Only if the breaker of the faulty section is not able to isolate this section, than the upstream breaker will trip.

A. Busbar differential protection

Differential relay operates on the differential current principle. The relay measures balance between the current supplied by incoming feeder E1 and summary of currents feeding the outgoing feeders A1, A2, A3. If the balance is not kept, the fault has occurred on busbar (Figure IV, situation F2).

Differential protection is the fastest type of power system protection, therefore is considered as primary protection of important busbars. Relay operates without time delay within shortest available time, usually 40-50ms. On the other hand the construction of this system is complicated and sensitive. Together with the requirements for detecting elements (current transformers) and input / output boards this solution is very expensive.

B. Overcurrent protection

Time definite overcurrent protective relay is the fundamental one, used as main or back-up relay for transmission and distribution systems of all voltage levels. The selective tripping of the fault (FIGURE V) with this protection is based on down-to-top current and time discrimination in positive way. It means that n+1 zone has higher setting value of both current and time delay. Current settings are based on the proper short circuit, motor starting, harmonic and load flow study with considering of motor starting peak loads and switching of high loads. In our case the basic setting of the feeder A1 protection is 0,3s that will ensure non-sensitivity of the relay on transient events and switching operations. In case the breaker will not operate due to any reason the fault will be cleared by upstream protection within the time of 0,9s.

Protection against the fault on busbar is ensured by second stage of incoming feeder protection with the time delay of 0,05s.

This commonly used principle, however, has a significant disadvantage in case of arc type of short circuit or earth fault. Under certain conditions the short circuit on busbar is not high enough to activate second stage of protective relay, but to high to cause significant damages up to the trip in 0,9s of first back-up stage. Communication of protective relays can significantly improve the protection feature in this case. Protective relay E1 will detect the faulty current and send request to downstream relay via GOOSE message (FIGURE VI). Return receipt will be received in less than 20ms. If the downstream relay doesn't detect the fault the upstream one can isolate the busbar even within less than 300ms required for downstream. Another possibility is publisher–receivers principle that is opposite case. Downstream relay will permanently publish the power circuit healthy signal to station bus. This signal will be received by upstream relay and in the moment of fault detection the upstream relay may immediately open the circuit. In this case the total time for operation is about 15-20ms shorter.

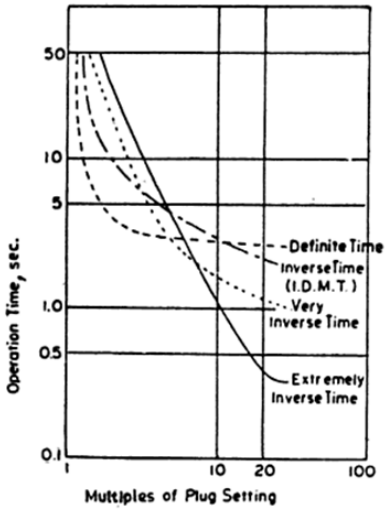


FIGURE V
OVERCURRENT PROTECTION CHARACTERISTIC

C. *Breaker failure*

Breaker failure protection is being used to improve the reliability of power system protection scheme. The operation principle is based either on self diagnostics of the respective protective relay or on the current detection on the upstream breaker. Within conventional power system protection schemes is this function performed by hardwired connection between different breakers. Specially in case of hugh power system this solution requires lot of wirings, means additional cost for engineering, installation and testing, but also risk for reliability and availability of whole system. IEC 61850 provides very comfortable solution for this problem by using GOOSE communication. Backup relays can mutually read all required information from the downstream relays and act the isolation of downstream faulty breaker. Breaker failure detection and isolation can be described as follows. Fault F1 is detected on feeder A1 by respective protective relay. Relay sends the tripping signal to the circuit breaker, however the breaker fails to open. Information about the tripping signal released together with the current values still measured (shall not be if the breaker is already off) are published on station bus and received by backup relay. This upstream relay (in our case relay on incoming feeder E1) will open related circuit breaker. Depending on priority this operation shall not take more than 20ms.

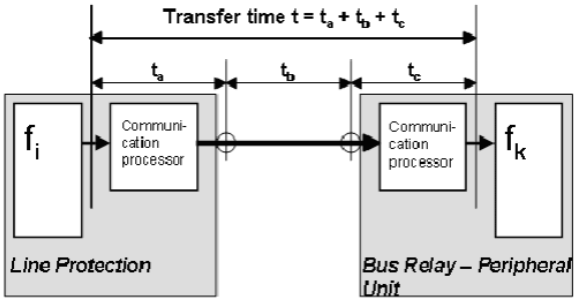


FIGURE V
PEER TO PEER COMMUNICATION

IV. Conclusion

The IEC 61850 as a huge communication protocol has a lot of advantages in the field of automation, control, signaling, safety and reliability of operation.

Regarding the contribution to the busbar protection within a substation the most valuable advantages are the reliable detection of the fault, optimizing of the tripping time and minimizing of hard wired connections within the substation.

Implementation the GOOSE messages containing current information, interlocking and intertripping signals enables clearing of the fault within second stage of the overcurrent protection 0,6 second faster than conventional approach. Implementation of the breaker failure protection is available without any additional wiring or design works using a message with the delay 3ms.

Contribution to the reliability of the system is easily implementable by change of network configuration. In our case the ring type network has several times higher level of reliability than the protection performed by hard-wired network.

References

[1] D. Tholomier - H. Grasset – M. Stockton, "Implementation Issues with IEC 61850 Based Substation Automation Systems " in Fifteenth National Power Systems Conference (NPSC). Bombay, 2008, pp. 473–478.

[2] S. B. Vicente and team, "New protection scheme based on IEC 61850". Zaragoza: CICRE. Available on internet: <http://www.aedie.org /11chlie-papers/220-Borroy.pdf >

[3] I. Mesmaeker, "How to use IEC 61850 in protection and automation". Available on internet: <http://www05.abb.com/global/scot/scot221.nsf /veritydisplay/fed26a71538479c3c12570d50034fbe4/$file/Rapport.pd>

[4] R. E. Mackiewitz, "Overview of IEC 61850 and Benefits". 2006. Available on internet: <http://morse.colorado.edu/~tlen5830/ho /Mackiewicz06IEC61850.pdf >

[5] D. Kececioglu, Reliability Engineering, Handbook. Lancaster: DEStech Publications, 2002, p. 535, ISBN 1-932078-01-0.

Testing of intelligent electronic device REF 543 using tester CMC 156

[1]*Miroslav KMEC (1st year),* [2]*Roman JAKUBČÁK (2nd year)*
Supervisor: [3]*Ľubomír BEŇA*

[1,2,3] Dept. of Electric Power Engineering, FEI TU of Košice, Slovak Republic

[1]miroslav.kmec@tuke.sk, [2]roman.jakubcak@tuke.sk, [3]lubomir.bena@tuke.sk

***Abstract* — This paper describes testing of intelligent electronic device (IED) REF 543. The aim of this article is to verify the overcurrent, undervoltage protection functions and their combination in different testing modules using operating software TEST UNIVERSE. The present paper is to point out the practical ways of using the above mentioned test equipment. The obtained results show the reliability and the quality of processed measured data and its need for usage during the testing actual complex digital protection systems.**

***Keywords* — CMC 156, REF 543, TEST UNIVERSE.**

I. Introduction

Protective relays now have many other functions besides protection. The advantages that modern IED provides over traditional relays are well documented. These advantages include fault location, event reports, and programmable logic that allow many functions to be included in one device, thus saving hardware and wiring costs.

The main task of IED is protecting networks or other equipment. To ensure the proper functioning of these networks it is necessary to use the appropriate type of IED. Faults conditions, as well as stable operating conditions are necessary to simulate, due to the reassurance that the system works reliably prior to commissioning. These tests should be repeated after a certain period due to the inactivity protective equipment under normal operating conditions. Repetition of these tests at regular intervals assures that the protection is working properly during its lifetime.

II. Feeder terminal REF 543

The REF 543 feeder terminal is designed to be used for protection, control, measurement and supervision of medium voltage networks. It can be used with different kinds of switchgear including single busbar, double busbar and duplex systems.

The protection functions also support different types of networks such as isolated neutral networks, resonant-earthed networks and partially earthed networks. In addition to protection, measurement, control, condition monitoring and general functions, the feeder terminal is provided with a large amount of PLC functions allowing several automation and sequence logic functions needed for substation automation to be integrated into one unit [1].

Fig. 1. Measuring and monitoring REF 543

This feeder terminal together with circuit breaker (Fig.1) is installed in switchboard which is situated at Department of Electric Power Engineering.

III. OMICRON CMC 156

This tester disposes solution for three-phase testing of IED. It is ideal for applications requiring a high degree of portability. CMC 156 (Fig.2) offers outstanding features and the absolute quality of its test sets. OMICRON has set new standards for advanced three-phase testing equipment in terms of flexibility, accuracy, portability and reliability [2].

For controlling CMC156, it is necessary to have installed software TEST UNIVERSE which offers a lots of test modules [3].

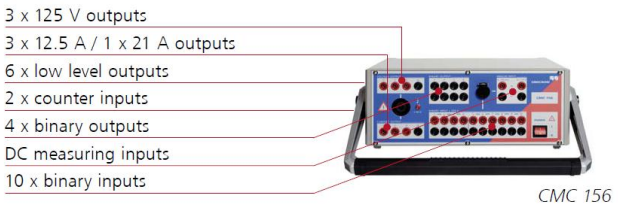


Fig. 2. OMICRON CMC 156[3]

IV. VERIFICATION OF OVERCURRENT FUNCTION

Overcurrent protection function is composed of two characteristics, namely [4]:

- IEC Normal Inverse (I>)
- Definite time (I>>)

These characteristics were tested in software TEST UNIVERSE, specifically in overcurrent module. The same parameters as for IED were entered into the module. Accurate configuration values of these characteristics are specified below in the Table I.

TABLE I
ELEMENTS - PHASE

Tripping characteristic	I Pick-up	Time	Reset Ratio	Direction
IEC Normal Inverse	0,90 I_{REF}	0,70	0,95	Non Direct.
IEC Definite time	8,20 I_{REF}	0,15 s	0,95	Non Direct.

After connecting the IED with CMC 156 and configuration of the settings of protective characteristics the test was run. The measurement result consists of individual points which were being tested and evaluated in table (Table II.) and a graphical form (Fig.3).

The test module compared set trip time of relay in accordance with permissible current (±50 mA) and time (±40 ms) tolerance, which was entered at the start of testing. This tolerance can be seen in the grey area in the vicinity of the test curve in Fig.3.

TABLE II
SHOT TEST RESULTS

Type	I_{SET}	t_{NOM}	t_{MIN}	t_{MAX}	Result
L1-1	0,50 A	No trip	No trip	No trip	Passed
L1-2	1,00 A	46,46 s	30,15 s	95,11 s	Passed
L1-3	1,50 A	9,54 s	8,27 s	11,14 s	Passed
L1-4	2,00 A	6,09 s	5,45 s	6,83 s	Passed
L1-5	2,50 A	4,75 s	4,30 s	5,25 s	Passed
L1-6	3,00 A	4,02 s	3,67 s	4,41 s	Passed
L1-7	3,50 A	3,56 s	3,26 s	3,89 s	Passed
L1-8	4,00 A	3,24 s	2,98 s	3,52 s	Passed
L1-9	4,50 A	3,00 s	2,76 s	3,25 s	Passed
L1-10	5,00 A	2,81 s	2,59 s	3,04 s	Passed
L1-11	5,50 A	2,66 s	2,46 s	2,87 s	Passed
L1-12	6,00 A	2,53 s	2,35 s	2,74 s	Passed
L1-13	6,50 A	2,43 s	2,25 s	2,62 s	Passed
L1-14	7,00 A	2,34 s	2,17 s	2,52 s	Passed
L1-15	7,50 A	2,26 s	2,10 s	2,44 s	Passed
L1-16	8,00 A	2,19 s	0,11 s	2,36 s	Passed
L1-17	8,50 A	0,15 s	0,11 s	2,29 s	Passed
L1-18	9,00 A	0,15 s	0,11 s	0,19 s	Passed
L1-19	9,50 A	0,15 s	0,11 s	0,19 s	Passed
L1-20	10,00 A	0,15 s	0,11 s	0,19 s	Passed

Individual test points were compared with tolerances and subsequently evaluated. Convenient points were marked in the table with the word Passed or in graphical form with green cross. All settings, tables, graphs and results of this module were generated into the protocol. At the end of protocol there was a general assessment of the overall test.

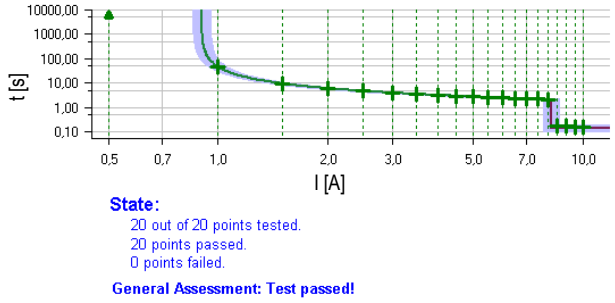


Fig. 3. Chart for fault type L1-E

V. VERIFICATION OF UNDERVOLTAGE FUNCTION

Three times voltage decrease below 0.85 U_N in six stages by ramping module was simulated in this test. Each decrease started from value 100 V and ended to 70 V with 100 mV step per 0.1 s (Table. III). Undervoltage function of feeder terminal was set at 85 % of U_N with 400 ms nominal time. The main objective of this test was to investigate the time and voltage values at pick-up and trip of the relay in consideration of set parameters.

TABLE III
RAMP STATES

	U_{L1-E}						
Ramp	From	To	Delta	Dt	d/dt	Steps	Time
Stage 1	100,0 V	70,00 V	-100 mV	100 ms	-1,0V/s	301	30,10 s
Stage 2	70,00 V	100,0 V	100 mV	100 ms	1,0V/s	301	30,10 s
Stage 3	100,0 V	70,00 V	-100 mV	100 ms	-1,0V/s	301	30,10 s
Stage 4	70,00 V	100,0 V	100 mV	100 ms	1,0V/s	301	30,10 s
Stage 5	100,0 V	70,00 V	-100 mV	100 ms	-1,0V/s	301	30,10 s
Stage 6	70,00 V	100,0 V	100 mV	100 ms	1,0V/s	301	30,10 s

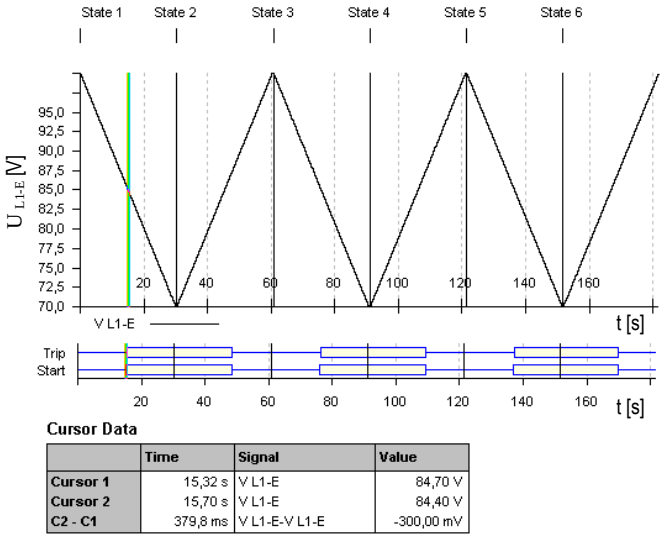


Fig. 4. Measurement results of undervoltage function

As it can be seen on the Fig. 4, the relay pick-up at 84.7 V and for time 379.8 ms tripped a simulated fault. It was similarly for the next two simulated voltage decrease when the time of trip did not exceed 400 ms. If we take the most accurate value of the voltage at the relay into account, the pick-up in this case would be 84.9 V. Aware of this idea, the total trip time of relay was extended for next 200 ms. We claim that the reaction of relay delayed for the established threshold about 200 mV what it is for time 200 ms. In practice, inaccuracy encounter interference relay does not validate. These inaccuracies are caused by several factors, which deal only with the producer himself. Since the CMC156 provides us a possibility of very accurate measurement it is possible to try such verification [5].

Actually, it depends on the time between pick-up and trip of relay. In our measurements this time ranged from 370 ms to 380 ms. We claim that our relay in all three tests worked reliably and trip simulated fault in a suitable tolerance.

VI. Combination of the Preceding Characteristics

In this measurement the combinations undervoltage and overcurrent protection function of REF 543 feeder terminal were tested with the use of module State Sequencer. This combination was simulated by eleven states as follows:

1. Pre-fault (t_{1max}= 300 ms)
2. Transition undervoltage (t_{2max}= 300 ms)
3. Normal state (t_{3max}= 300 ms)
4. Transition overcurrent (t_{4max}= 150 ms)
5. Normal state (t_{5max}= 300 ms)
6. Undervoltage (t_{6max}= 500 ms)
7. Normal state (t_{7max}= 300 ms)
8. Overcurrent (t_{8max}= 500 ms)
9. Normal state (t_{9max}= 300 ms)
10. Overcurrent + undervoltage (t_{10max}= 500 ms)
11. Definitely trip (t_{11max}= 200 ms)

An exact duration of time is specified for each state. Detailed settings are shown in Table IV.

TABLE IV
TEST SETTINGS OF MODULE STATE SEQUENCER

STATE	1.	2.	3.	4.	5.	6.
V L1-E	100V	70 V	100V	100 V	100V	70 V
	0,00 °	0,00 °	0,00 °	0,00 °	0,00 °	0,00 °
	50 Hz	50 Hz	50 Hz	50 Hz	50 Hz	50 Hz
I L1	1A	1A	1 A	8 A	1 A	1A
	0,00 °	0,00 °	0,00 °	0,00 °	0,00 °	0,00 °
	50Hz	50Hz	50Hz	50 Hz	50Hz	50 Hz
Max.Time	300ms	300ms	300ms	150ms	300ms	500ms

STATE	7.	8.	9.	10.	11.
V L1-E	100 V	100 V	100 V	70 V	0V
	0,00 °	0,00 °	0,00 °	0,00 °	0,00 °
	50Hz	50Hz	50 Hz	50 Hz	50Hz
I L1	1A	8 A	1 A	8 A	0A
	0,00 °	0,00 °	0,00 °	0,00 °	0,00 °
	50Hz	50 Hz	50 Hz	50 Hz	50Hz
Max.Time	300ms	500ms	300ms	500 ms	200ms

Overcurrent and undervoltage protection function was set according to the values specified in Table V. The main objective of this test was measuring trip time of relay in depending on the type and duration of fault [6].

TABLE V
TEST SETTINGS OF FEEDER TERMINAL

Tripping characteristic	Value	Time	Direction
Overcurrent – Definite time	6,5Iref	0,20 s	Non Direct.
Undervoltage – Definite time	0,85Uref	0,40 s	Non Direct.
Iref = 1A; **Uref** = 100V			

In the second and fourth state there were transitions faults simulated of which the time set was shorter than the relay operate time. In these conditions, we can see from Fig.5, only the start of relay without its trip. In the sixth and eight states there were continuous faults simulated where the duration of the fault is longer than set trip time of relay. In this case, the relay tripped at the time according to the type of fault. The tenth state was combination of undervoltage and overcurrent characteristic, where we can see the joint start both functions and finally trip only the overcurrent function due to its shorter duration of action.

For a more detailed analysis, it was found that the relay seen earlier emergence overcurrent (30 ms latency) as an undervoltage (40 – 50 ms latency), while both faults arise together. From the time when relay started, trip time for each function was in the range:

- Undervoltage (370 – 380 ms)
- Overcurrent(180 – 190 ms)

In neither of cases exceed trip time of relay was expected. Feeder terminal worked reliably for various transitional and continues faults as well as their combinations. We argue that terminal REF 543 in all states operated reliably and trip simulated fault in a suitable tolerance.

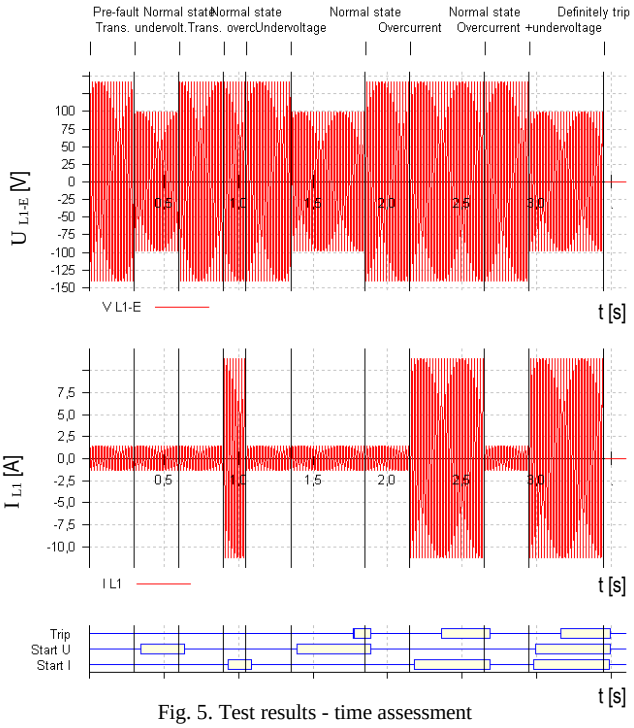


Fig. 5. Test results - time assessment

VII. Conclusion

This article points out the multilateral use of OMICRON CMC 156 test equipment and its operating software Test Universe. The aim of this article is to describe the practical ways of using the above mentioned test equipment. The obtained results show the reliability and the quality of processed measured data and its need for usage during the testing actual and future complex digital protection systems.

References

[1] ABB, Product guide, Protection and control, Feeder terminal REF 543. [Online].< http://www.abb.com/product/db0003db004281/c12573e700330419c2256e6300300dd6.aspx?productLanguage=us&country=00>

[2] Test and Measurement Hire Ltd, OMICRON CMC 156 [Online]. < http://www.instruments4hire.co.uk/test-equipment/product-71-OMICRON/OMICRON-CMC-156--12-5A-/>

[3] OMICRON, Test Universe software, PC software suite for advanced secondary testing. [Online]. < http://www.omicron.at/en/products/pro/secondary-testing-calibration/sw/ >

[4] CHLADNÝ, Vladimír: Ochrany v elektrizačných sústavách, návody na cvičenia. Košice. 1992. ISBN 80-7099-133-X

[5] JANÍČEK, František – CHLADNÝ, Vladimír – BELÁŇ, Anton – ELESCHOVÁ, Žaneta: Digitálne ochrany v elektrizačnej sústave. Bratislava. 2004. ISBN 80-227-2135-2

[6] CHLADNÝ, Vladimír: Digitálne ochrany v ES. Košice. 2007. ISBN 978-80-8073-798-6

The Dependence Aging Factor of Oil Insulation of Temperature.

[1]*Vieroslava SKLENÁROVÁ (3st year),* [2]*Pavol HOCKO (2st year)*
Supervisor: [3]*Roman CIMBALA*

[1]Dept. of Dept. of Electric Power Engineering, FEI TU of Košice, Slovak Republic

[2]Dept. of Dept. of Electric Power Engineering, FEI TU of Košice, Slovak Republic

[3]Dept. of Dept. of Electric Power Engineering, FEI TU of Košice, Slovak Republic

[1]vieroslava.sklenarova@tuke.sk, [2]pavol.hocko@tuke.sk, [3]roman.cimbala@tuke.sk

***Abstract*— This article deals with problem of degradation process in the oil. Attention is devoted to the oil insulation system and coefficient of aging for oil insulation system through value of IRC Analysis in dependence on the temperature.**

***Keywords* — Information, aging, insulation system, temperature, transformer oil.**

I. INTRODUCTION

Economics and technique would be too interconnected. Economic theory is to apply natural to techniques practice and conversely. Corporate databases contain a wealth of information, but sometimes it is a problem with their proper completing. Database files contain millions to billions of different segmentation and organized records. The problem will be information overload, understanding the value of information.

Imagine different models cases:

Situation 1: Responsible employees of the electric power plant are deciding whether the transformer can be operated or must be exchanged for a new one. They know all the technical information of transformers that have been replaced.

Situation 2: It is necessary to decide on the basis of different criteria, which equipment must be diagnosed in order to prevent disturbances

Situation 3: Oil filling of transformer has unsuitable parameters. It is necessary to decide whether or not dried, regenerated or replaced.

In these situations can be easier to find solutions through the Business Intelligence systems.

Business Intelligence is a complex process of transforming data into information and converts that information to knowledge through exploration. At first glance it might seem that between data and information can be placed equal sign. However, dates have become the information only if [1]:

- you have data
- you know that you have data
- you know where the data are stored
- you have access to the data
- you can trust to source of data

Business Intelligence can provide valuable information for decision support. Business Intelligence is defined as a set of drafts and methodologies that optimize decision-making systems to use systems based on the metrics. The role of business intelligence is to transform data into information and convert that information to knowledge. For this reason, Business Intelligence can be utilized corporate systems, as well as in laboratories. One of the ways in which it can be applied is aging transformer oil.

II. AGING OF TRANSFORMERS

Aging is a set of processes that the operating conditions lead to changes in the physical, chemical and electrical properties, which will cause the operational safety devices in terms of insulation properties.

Aging of transformers is reflected adversely primarily for electro insulation systems that are created potential barriers. The violation of potential barriers is the immediate cause of the failure. Electrical insulating system of power transformers is formed of dielectric oil-paper.

Life of the transformer is limited mainly paper insulation life, where mechanical strength is gradually decreasing. Life of paper insulation is usually significantly shorter than that of other structural elements. Disproportion between the life of paper insulation and other structural elements is considerably larger, if is not received treatment of oil filling of transformer. The aging between oil and paper is reciprocal dependence. Products of oil decomposition are absorbed by the paper and conversely, which is accelerated the aging process. If oil filling of transformer is not replaced (or regenerated) before the end-of-life than is accelerated the irreversible degradation of the paper insulation, and thus is accelerated the loss-of-life transformers too.

The aging oil is caused by others influence. Effects can be divided according [2]:

- chemical substances of acid and alkaline nature, and also catalysts and water,
- oxygen, together the action of heat or other energy,
- energy (thermal, electric field and electric discharges),
- uniformity of load, load levels, vibration and cooling efficiency.

Operational safety is decreased mainly for the following reasons:

- deterioration of cooling as a result aging sludge deposits in oil ducts and the winding,
- reduction the mechanical strength of cellulose insulators due to attack of acid incurred by aging,
- reduction electric strength of insulating oil and oil-paper system due to aging products, in particular water.

Water, heat, and oxygen are the catalyst, the accelerator, and the active reagent in the oxidation of the oil in oil-filled transformers.

III. OIL AS DIAGNOSTIC MEDIUM

The solid component insulation system is difficult available. Of course, sampling of insulation is not possible for analysis during transformer operation. For study of properties of insulation is therefore necessary to use indirect methods.

The changes in the solid insulation can be monitored with analyzing liquid component, which is an excellent carrier of diagnostic information.

This is mainly because the deterioration of the solid component are secreted gases in severe operating conditions of temperature, high-voltage failures or partial discharges, which are partially dissolved in the oil. Their analysis can determine the type and severity of disorder and its causes. It must be borne in mind that these products can be secreted by the aging of oil too, so there may be a distortion of the results.

Isothermic relaxation current analysis (IRC analysis) is enabled to analyze non-destructive aging of system. IRC analysis is allowed to define the elements of equivalent circuit and calculate the dielectric constant of aging A. It is necessary to take into account the time zone and the relevant spectrum, while the equivalent circuit elements must be independent of voltage and frequency.

Coefficient of aging A has been empirically determined to assess and quantify the replenishment [3].

If $\tau_3 \phi \mathbf{3}.\tau_2$

and $\tau_2 \phi \mathbf{3}.\tau_1$

then to coefficient of aging A apply:

$$A = \frac{Q(t_3)}{Q(t_2)} = \frac{1 + \frac{a_2.\tau_2}{a_1.\tau_1}.(1 - e^{\frac{\tau_1}{\tau_2}}) + \frac{a_3.\tau_3}{a_1.\tau_1}.(1 - \frac{1}{e})}{1 + \frac{a_2.\tau_2}{a_1.\tau_1}.(1 - \frac{1}{e}) + \frac{a_3.\tau_3}{a_1.\tau_1}.(1 - e^{\frac{\tau_1}{\tau_2}})} \qquad (1)$$

where: a_i, τ_i - amplitudes of elementary streams and equivalent times in the i-th branch of the replacement model of the dielectric in the development of three RC members (obtained by calculation in the third approximation).

In our measurements, we showed the impact of atmospheric moisture. We were looking for a way to eliminate this influence

We made two repeated measurements of new mineral oil MN by method of IRC analysis. The temperature is increasing gradually by step 10° C and monitoring the polarization processes the during the 1000 s.

The first day we made measurements at 20 to 50° C, the second day measurements took place at 60 to 80°C. We made the measurements at 90 and 100° C last day. Temperature of oil was allowed to stabilize around 1 hour before measurement.

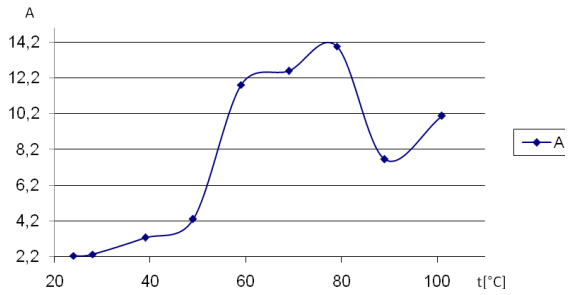
Subsequently, we evaluated value of IRC analysis and approximated to third polarization processes. Next we calculated the aging factor A (Table I) during relationship (1).

TABLE I
CALCULATE THE AGING FACTOR A FOR MN

	MN1-24	MN1-30	MN1-40	MN1-50	MN1-60	MN1-72	MN1-81	MN1-91	MN1-101
	24	**30**	**40**	**50**	**60**	**72**	**81**	**91**	**101**
I_{m311}	48638	48669	55040	74930	112510	150389	35545	110677	81811
I_{m321}	6045	6323	8767	7638	6684	7917	10524	20911	30845
I_{m331}	1275	995	1206	1569	1377	2693	13644	15359	25545
I_{m301}	498	616	1061	1654	1862	2831	3979	8852	12111
τ_{311}	0,31	0,39	0.1	0,46	0,43	0,37	1,34	0,44	0,77
τ_{321}	3,4	5,28	5.1	5,48	8,34	8,7	39,1	17,83	16,88
τ_{331}	46,2	85,3	160,7	224,5	1169,5	832,6	1296,0	391,6	472,7
Q_{t3}	4,83	5,59	7,89	8,67	23,19	27,71	244,31	86,72	130,44
Q_{t2}	2,14	2,38	2,41	2,01	1,96	2,20	17,49	11,34	12,95
A	**2,26**	**2,35**	**3,28**	**4,31**	**11,80**	**12,59**	**13,97**	**7,65**	**10,07**

The figure (Fig.1) shows the dependence coefficient of aging A of temperature.

Fig. 1 Coefficient of aging A of MN



Second measurements on this sample monitoring the polarization processes during the 1000 s and the temperature is increasing gradually by step 10° C too. The first day we made measurements at 20 to 50° C, the second day measurements took place at 60 to 90° C. Oil temperature was allowed to stabilize around 6 hours before measurement.

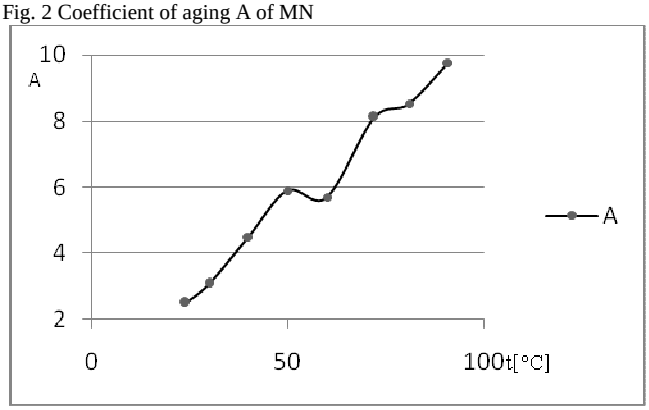
For second measurement we calculated the aging factor A (Table II) during relationship (1) too.

TABLE II
CALCULATE THE AGING FACTOR A FOR MN

	MN-24	MN-30	MN-40	MN-50	MN-60	MN-72	MN-81	MN-91
	24	30	40	50	60	72	81	91
I_{m311}	60437	77176	52518	136419	65310	45863	134393	92209
I_{m321}	7924	10227	4732	6050	8035	11073	19084	30203
I_{m331}	1043	1268	2121	3073	5478	9976	14465	24491
I_{m301}	870	1265	1025	2939	3301	5035	7714	11101
τ_{311}	0,38	0,37	0,8	0,4	0,69	0,92	0,42	0,7
τ_{321}	5,53	4,73	13,86	12,46	23,05	25,55	19,51	18,28
τ_{331}	111,67	153,66	295,85	357,58	402,14	553,2	482,47	497,75
Q_{t3}	6,1138047	7,00721	12,0019	15,1107	36,01072	90,38278	85,7524	128,938
Q_{t2}	2,4511262	2,27768	2,67031	2,56286	6,321118	11,1419	10,0697	13,2172
A	**2,494284**	**3,0765**	**4,4946**	**5,896**	**5,69689**	**8,11197**	**8,5159**	**9,7553**

The figure (Fig.2) shows the dependence coefficient of aging A of temperature for second measurement.

Fig. 2 Coefficient of aging A of MN



Our measurement results were affected, as is shown in Figure1and Figure 2

By comparing the measurements we see that in both of these measurements again showed effects after discontinuation of atmospheric moisture measurements. Lower effects are manifested with prolonged heating the oil in the second measurement.

IV. CONCLUSION

Water is a natural part of oil. In our laboratory conditions is difficult to ensure the elimination of access to atmospheric moisture. In our opinion, the measurement is occurred of access to atmospheric moisture to the sample discontinuation of. If the measurement sample of oil is gradually heated, then moisture getting rid continuation in part again. Comparing the two measurements, we see that the second measurement again showed effects of atmospheric moisture in break of measuring. Our measurements are significantly demonstrated affects of moisture the measured results, and there by evaluating the coefficient of aging A. Because our experiments are time consuming it is impossible to a continuous increasing temperature. We suggest the measurement with long-term heating oil.

ACKNOWLEDGMENT

This paper was developed within the Project "Centrum excelentnosti integrovaného výskumu a využitia progresívnych materiálov a technológií v oblasti automobilovej elektroniky", ITMS 26220120055 and wish to acknowledge Scientific Grant Agency of the Ministry of Slovak Republic and Slovak Academy of Science for funding of experimental works in the frame of VEGA No. 1/0487/12 grant.

REFERENCES

[1] Ľ. Lacko, "Je váš biznis inteligentný," in *Reseller Magazine*, číslo1, ročník 10. MK ČR E 14482, ISSN 1214-3146©DCD Publishing, s.r.o., pp. 23.

[2] V. Barborka,: "Diagnostika transformátorových olejů v návaznosti na prodloužení životnosti transformátorů," in *Elektro odborný časopis pre elektrotechniku*, číslo 07, ročník 2002, dostupný: http://www.odbornecasopisy.cz/index.php?id_document=25085

[3] R. Cimbala, –I. KRŠŇÁK, – I. KOLCUNOVÁ, : The Computation of Influence of Steady Element on Polarization Spektrum, Journal Acta Polytechnica Prague, Vol. 43, No. 2/2003, ISSN 1210-2709, Prague.

The Impact of Photovoltaic Power Plant on the Power System

[1]*Jozef DUDIAK (2nd year Ing.),* [2]*Pavol HOCKO (3rd year PhD.)*
Supervisor: [3]*Michal KOLCUN*

[1]Dept. of Electric Power Engineering, FEI TU of Košice, Slovak Republic
[2,3]Dept. of Electric Power Engineering, FEI TU of Košice, Slovak Republic

[1]jozef.dudiak@student.tuke.sk, [2]pavol.hocko@tuke.sk, [3]michal.kolcun@tuke.sk

***Abstract*—this paper describes control of the power system with installed photovoltaic power plants. Photovoltaic power plants are connected to the power system with conventional power plants. The power system consist of two separate areas which are connected together with two inter- tie lines. There is the simulation of photovoltaic power plant during cloudy and clear sky. The difference between summer and winter supply from this power plant is also compared as well. That simulation is made in simulating program Modes. Subsequently, we compare the produce of power plants with installed photovoltaic and also without it. The second part of this paper discuss about frequency deviation caused by variable supply of active power from photovoltaic power plant.**

***Keywords*—photovoltaic power plant, support services, frequency deviation, MODES.**

I. INTRODUCTION

Photovoltaic power plants are characterized by variable supply of active power, which depends on the intensity of sunlight. The requirements of the European Union to the 2020, to produce 20% of electricity from renewable energy sources established the trend of installing large amounts of PV. If the proportion of energy mix in power resources exceeds a certain level, it would increase problems with maintaining of power balance and therefore the reliability of electricity supply will be threatened.

The reliability of power supply is defined as its safety and appropriateness. Safety is the ability of the power system to resist faults, short circuits or failures of devices and safety is related to dynamic stability. The adequacy is determined by the ability of the power system to ensure power supply for customers, i.e. -maintain a power balance at any time.

Large amount of installed PV power plants in the power system threatens the adequacy of the power system. This is in case, when other sources with provided performance aren´t enough to compensate the variable production from PV. Power balance have to be maintained in every interconnected power system, so that power of resources must cover consumption, losses and also planned exchange between others regulatory areas. Operating rules establish deviations from the planned exchanges that wouldn´t be exceeded during normal operation.

II. MODES, THE MEANS FOR MODELING IMPACT OF PHOTOVOLTAIC POWER PLANTS

A. Outgoing circuit

The power system of IEEE39 is interpreted as the electricity grid region New England. It consists of 39 nodes, that are distributed into three voltage levels and 10 generators that operate in our case in two areas.

Generator *GEN1*, which works in the area No. 2 represents group of several generators in a single unit. This generator due to high output power works directly to the node 39, which is on voltage level of 400 kV. All other generators, connected to the power system works to base area No. 1. These generators are working to the nodes on voltage level 22 kV, where the voltage is transformed by using transformers 22/400kV.

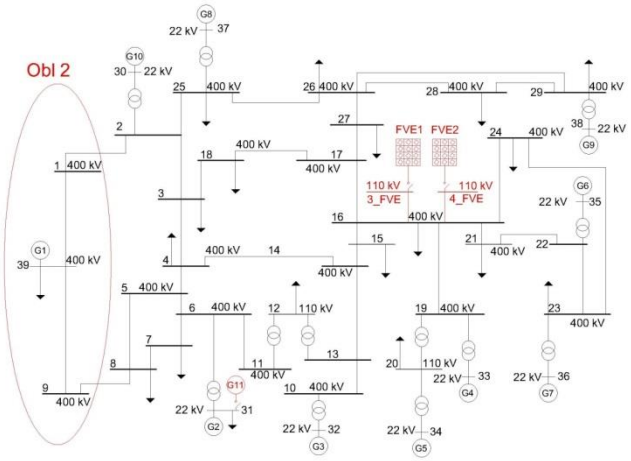


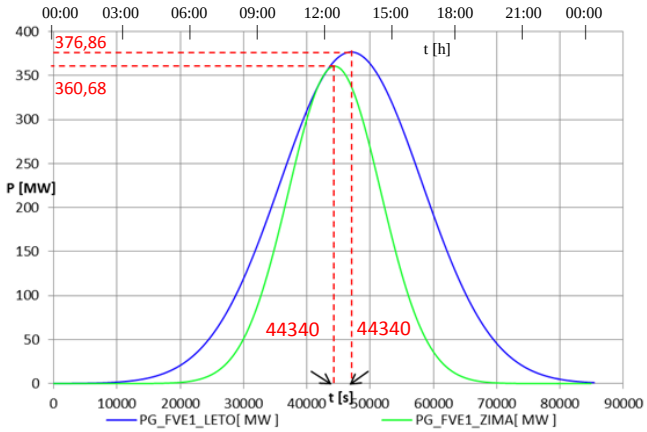
Fig. 1 Initial scheme for modeling photovoltaic power plants

Node 3_FVE and 4_FVE are added to modified power grid. Nodes are on 110 kV voltage level and they are used for connecting PV power plant.

B. Modeling of clear sky during summer and winter day

For modeling impact from PV power plants on power system we have created a new case in which we set *NODE3_FVE* and *NODE4_FVE* as active nodes and also transformers between these nodes and node *NODE16* as active too.

In this case, we also set solar power plants as active and in the case of calculation we set information about day time to value 0. It means that the operation of the long-term dynamic simulation will start from the time 0 (midnight of the day). For the PV power plants that are connected to the power system we set the type of source parameters to *LETOJ* for summer day and *ZIMAJ* for winter day. The installed capacity in FVE1 is 550 MW and in FVE2 is 150MW.



PG_FVE [MW] – Active power from power plant

Fig. 2 Process of active power from photovoltaic power plant.

Real power production of PV is shown in Fig. 2. During the winter sunny day PV power plant supply less active power than during summer sunny day. The maximum of active power during winter day is 360,68 MW at the time of 12:32 am and the maximum during summer day is 376,86 MW at the time of 13:05 pm. Also we can see that during winter days we have later sunrises and earlier sunsets, that affects the total amount of power supplied from PV power plant during winters. According to Fig. 2 we can note that persistence of maximum power supply from PV is significantly shorter in winter.

Calculation of energy supplied from photovoltaic power plant

To compare total electric power supplied by photovoltaic power plants, we have to calculate the definite integral from area defined by supplied active power curve and timeline. Considering we do not know analytical prescription of the function f(x), but the function is defined as table. We also know the limits of the integral < a;b >, so we choose trapezoidal method for solving definite integrals.

$$\int_a^b f(x)dx \approx \sum_{n=1}^{n} \frac{f(x_{i-1}) + f(x_i)}{2}(x_i - x_{i-1}) \qquad (1)$$

When we start to integrate this area according to (1) the result of the definite integral will be area of each trapezoid. Counting these areas together we get the total energy supplied by photovoltaic power plant FVE1 throughout the whole day.

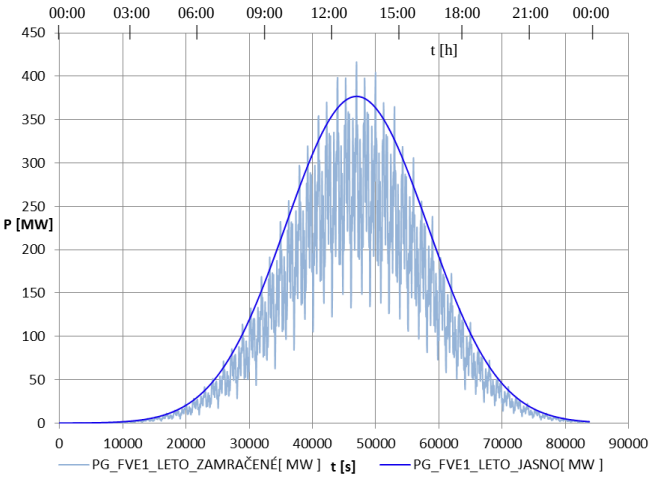
TABLE I
COMPARISON OF SUPPLIED ACTIVE POWER

Electric energy	CLEAR	CLOUDY	Difference	Difference in %
SUMMER	2936,35 MWh	2133,26 MWh	803,09 MWh	27,36
WINTER	1826,31 MWh	1357,83 MWh	468,48 MWh	25,65
Difference	1110,63 MWh	775,43 MWh		
Difference in %	37,82	36,35		

For clear summer day we get the total supplied electric power of 2936.95 MWh. For a clear winter day, the value of electric power from photovoltaic power plant is 1826.31 MWh that is 62.18% from the total energy supplied during the summer day. From this calculation, we can show that photovoltaic power plant with installed capacity of 550 MW supply during winter clear days about 1110.63 MWh less than during the summer clear days, so that is about 37.82% less.

C. Modeling of cloudy sky during summer and winter day

Modeling of cloudy day we made on the same photovoltaic power plants with installed capacity of 550 MW for FVE1 and 150 MW for FVE2. Model parameters of photovoltaic power plant, we reset source parameters to SLETO1 for summer day and SZIMO1 for modeling winter day. Results of dynamic simulation for summer day are shown in Fig. 3. From this course it seems that during cloudy days the supplied active power is much more unstable, as during clear days. These deviations in the production from PV power plants may cause power system fluctuations or in some specific cases blackout of the power system. With large amount of installed power in PV, it is difficult to maintain a sufficient margin for balancing these changes.

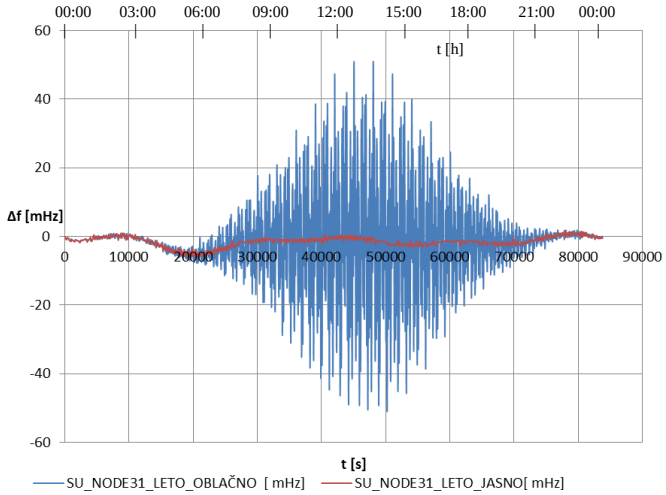


PG_FVE [MW] – Active power from power plant

Fig. 3 Process of active power from photovoltaic power plant during cloudy summer day

By comparing the curves during the clear and cloudy day, we can see that cloud causes the variable supply of active power. When we calculate the definite integral from curve of power supply, we find out that during the cloudy day FVE1 supply in total 2133.26 MWh of electric energy in to the power system. Compared with the clear summer day it is about 27.36 % less of electric energy.

Progress of the frequency deviation from the nominal value curve is shown in Fig. 4. Clear day weather conditions doesn´t produce the frequency deviations from the nominal value, small frequency deviations are automatically controlled by action of secondary regulation of active power, respectively. Active power supplied from PV during clear day leads to a reduction of power supplied from other conventional sources. During cloudy days there is a frequency deviation, due to speed of changing performance from PV and slower response from other conventional sources to resulting deviation.

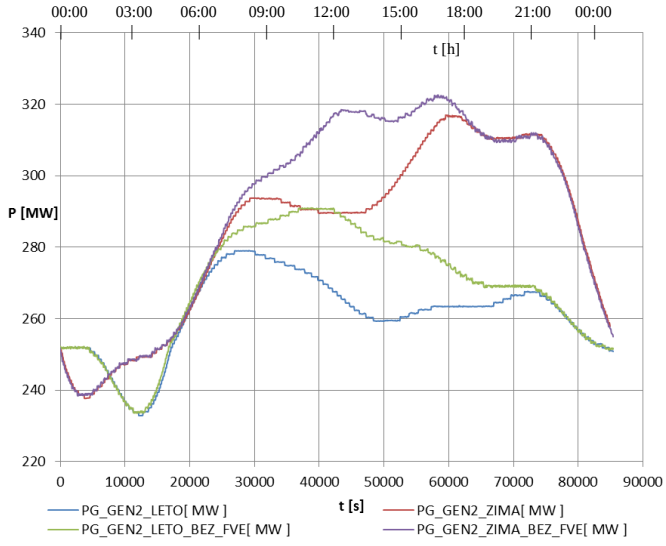


SU [mHz] – Frequency deviation from nominal value

Fig. 4 Process of frequency deviation with and without FVE

The impact of photovoltaic power plant

On the following figure Fig. 5, we can see the impact of PV on other conventional power plants. Unit *GEN2* automatically reduces supplied active power to the power system as the effect of secondary regulation process. Electric energy produced from photovoltaic power plants should be bought with priority, so other conventional sources must regulate that change in production during the day.



PG_FVE [MW] – Active power from power plant

Fig. 5 Power supply of GEN2 with and without installed photovoltaic power plant.

From this process is obvious that photovoltaic power plant supplies active power to the power system at the time interval of maximum load in the system, according to the daily load diagram. From this perspective, this is quite advantage of PV. But in real power system the power supplied from PV won't be that smooth as it was in our simulations.

III. Conclusion

Photovoltaic power plants are during clear days less problematic because nowadays we already have quite reliable means of weather forecast that ensure nearly exact prediction of weather and then predict of active power supply to the power system from PV. Despite the high quality of technology these renewable energy sources are considered as difficult to predict. In case of cloudy days is the power supply variations that can cause problems in maintaining stability. The great advantage of photovoltaic power plant is also that they supply active power to the system during the day, when the energy consumption is on the highest level.

In every case we need to have total installed capacity of PV in 100% reserve in the form of tertiary regulation on conventional sources.

References

[1] DUDIAK, Jozef.: Modelovanie podporných služieb v programe Modes. Košice: TU, 2013. Diplomová práca.

[2] KOLCUN, Michal et al.: Prevádzka elektrizačnej sústavy. Košice: TU, 2007. 306 s. ISBN 978-80-8073-837-2.

[3] MASLO, Karol.: Popis modelování přechodných dějů v ES programem MODES v.2.2./8.

[4] MASLO, K.: Uživatelská příručka programu MODES 2.2/12.

[5] Kučerová, A.:Numerická integrace lichoběžníkovou metodou[online]. Praha: CVUT, SFV, 2012 [citované 5. 12. 2012]. Dostupné na internete: <http://klobouk.fsv.cvut.cz/~anicka/teaching/ypv1/priklady/numint.pdf>

[6] MASLO, K.: Maintaining power balance in a system with renewable energy sources.

[7] MASLO, K.: Long – term dynamic modeling of renewable energy sources.

The Influence of PSS Parameters to Transient Stability of Power System

[1]*Matúš Novák (2nd year)*, [2]*Zsolt ČONKA (1st year)*
Supervisor: [3]*Michal KOLCUN*

[1,2,3] Deptartment of Electric Power Engineering, FEI TU of Košice, Slovak Republic

[1]matus.novak@tuke.sk, [2]zsolt.conka@tuke.sk, [3]michal.kolcun@tuke.sk

***Abstract*— This paper deals with analysis of behavior of electric generators in power system under various operating conditions. Authors examine influence of various configurations of generator excitation systems to the transient stability of power system, and also influence of excitation system parameters. Investigations proved that the selection of convenient parameters could help to improve efficiency of regulation of excitation systems and damping of generator oscillations.**

***Keywords* - electric power system, stability, automatic devices, transient stability, power system stabilizer, PSLF, oscillation damping.**

I. INTRODUCTION

At the present moment, electrical utilities must raise their effort to achieve their main goal – ensure reliable supply of electricity to their customers, in the required quality and quantity, without unwanted interruptions. It is due to opening electricity market, which results in tightening the laws and regulations. It also must be taken into account increasing consumption of electricity, and a growing number of local sources of electric power. Power systems, built mostly in the last century, when such trading of electricity, as today, was not known, became therefore even more complicated. In power systems of these days, the smallest mistake during operation can bring big failures. These failures can involve large economic losses, not only for consumers, but especially for the power system operator, who must pay a fine for non-delivering of the energy.[3]

Rotating electric power generators are very important part of power systems, because they still produces largest part of all consumpted energy. In their operation, many factors must be taken into account, for example reliability, safety, economical efficiency of their operation. When the behavior of the power generator is going to be examined, not only machine should be considered, but also all devices related to machine, such as turbine, governor, exciter, exciter regulator, , automatic voltage regulator (AVR), power system stabilizer (PSS), limiters and protection devices, must be taken into account.

This paper brings the examples of different configurations of excitation systems and their influence to the power system stability.

II. EXCITATION SYSTEMS

The main, and most important function of excitation system, is to provide direct current to the synchronous machine field winding. Excitation systems can have, and does usually have other functions, which are necessary for proper function of power system, such as control and protective functions, which influences field voltage. Protective functions are aimed to ensure that the no capability limits of generator would be exceeded, for example, underexcitation limit. Purpose of control functions, is to control generator voltage and reactive power, and if possible, to act to enhance power system stability.

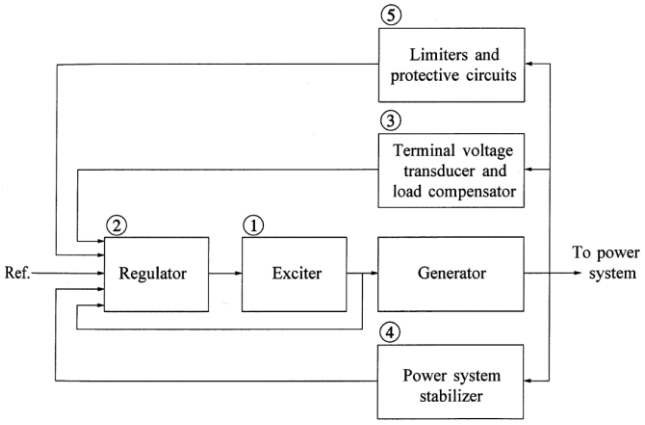


Fig. 1 Block diagram of a synchronous generator excitation control system. [1]

Following is a brief description of various subsystems of excitation system, which can be identified at the picture.

A. Exciter

Exciter acts as a power source of DC power to the field winding of generator. Over years, excitation systems have taken different forms, but they can be generally classified into three categories, based on the principle of dc power source.

DC exciter is a first, oldest principle. It uses dc generator to supply dc current to generator field winding directly through the slip rings. Voltage regulators for this type of systems uses different principles, such as continuous rheostatic types, or rotating amplifiers. DC excitation systems are gradually disappearing, but some are still in service, although with voltage regulators replaced by modern, solid state electronic devices.

AC *exciter is* a second, newer principle. It uses ac generator connected to the field winding through diode rectifier. Rectifier can be non-controlled, when diodes are used, or controlled, when thyristors are used. Both types supply dc current to field winding through slip rings. There is also a special type called *brushless excitation system*, which has higher reliability due to missing slip rings, because it has rectifier diodes mounted to shaft between AC exciter and field winding of generator. This type was designed for high power generators, which have high field currents – for example, 600 MW generators needs 1 MW of dc power supplied to field winding.

Static exciter is a third, newest principle. It is called stationary due to fact, that all components of this system are stationary. Static exciters became more distinct with the development of power semiconductor devices. Most of new generators have this type of exciters. Static exciter consists from power supply and from controlled or uncontrolled rectifiers, from which is DC power supplied through slip rings to field winding. There are three main categories of static exciters, based on the power source for exciter, and there are for example exciters fed from auxiliary bus, or from saturable transformer connected beetween armature winding of generator and neutral grounding.

B. Regulator

Regulator works as processor and amplifier for input signals, to prepare them to the appropriate level and form for control of exciter, including regulation and stabilization of exciter. Voltage regulator is sometimes referred as Automatic Voltage Regulator (AVR).

Main function is to maintain the generator stator voltage, through which other values can be affected, such as reactive power. There are usually two regulators, AC regulator serves as main, and through this auxiliary control and protective functions acts, and DC regulators, which hold constant field voltage and is referred as manual control. In regulator are also included excitation system stabilizing circuits, which affects only stabilizer, not to be confused with power system stabilizer (PSS), and also load compensation circuitry.

C. Limiters and protective circuits

Limiters and protective devices include many protective and control functions. Their purpose is to ensure that the any capability limit of any device would not be exceeded. [1]

Limiters include underexcitation limiter (prevents loosing steady-state stability and overheating of generator stator core), overexcitation limiter (protects from overheating due to prolonged field overcurrent), Volts-Hertz limiter (prevents the excesive magnetic flux resulting from overvoltage and low frequency) and also field shorting circuits, which serves in some conditions as path for negative field current induced in field winding, by bypassing the exciter, and can have form of crowbar or a varistor.

D. Terminal voltage transducer

This device senses generator terminal voltage, rectifies it and compares it with reference voltage, and passes it to the voltage regulator (AVR).

E. Power system stabiliser

Power system stabilizer (PSS) was brought into excitation systems to improve stability of power system, by damping of oscillations induced mostly by short circuits in system. It is possible by using auxiliary stabilizing signals. Most commonly used signals are frequency deviation, shaft speed, accelerating power.

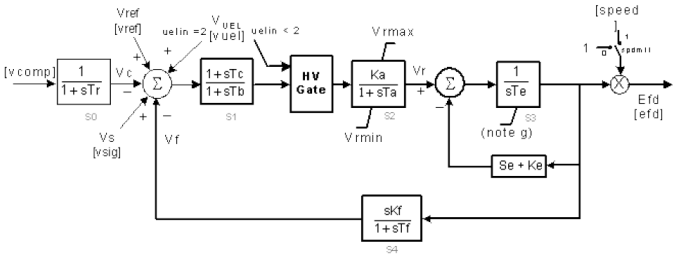


Fig. 2 Block structure of DC1A excitation system model, used in PSLF.[4]

All of these components should be properly modeled, when analyzing power system transient stability, especially exciter, regulator, power system stabilizer, to obtain valid results. Usually it is inadequate to use very detailed model of excitation system, such detail is often considered too great for general system studies. There are usually used some techniques to obtain simplified model, which is more practical to type of study, for which it is intended. Using these techniques, the appropriate parameters of reduced models, as gain and phase characteristics are selected, so the models match detailed model in frequency range usually from 0 to 3 Hz. Also all significant nonlinearities are considered. [2][5]

During the years of research, appropriate structure of different reduced models was standardized by IEEE, to represent wide variety of excitation systems currently in use. Example of such model, slightly modified to use in the PSLF software, is shown on a Fig. 2. This model is extended with optional speed multiplier. This model does not include PSS.

In next chapter, modeling of power system, including complete excitation system will be shown.

III. POWER SYSTEM MODEL

For examination of PSS parameters influence into transient stability, a small power system consisting of 11 nodes and three generators, will be used, similarly as in [1]. Power system is shown on Fig. 3.

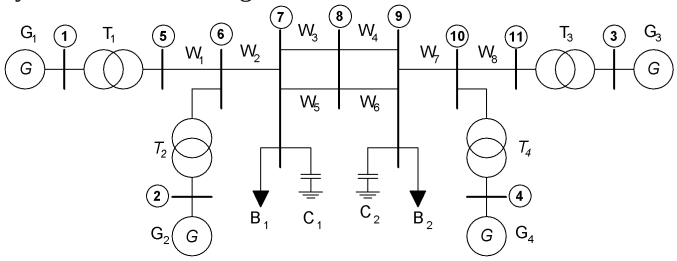


Fig. 3 11-bus, 4-generator power system

All generators has a 900 MVA rating on the 20 kV, with step-up transformers with same rating and 20/230 kV base. Generators are supplying power to the load at bus 7 – S_7 = (967+j100) MVA and bus 9 - S_7 = (1767+j100) MVA. There are also shunt capacitors connected at these buses, with supplied reactive power of Q_{7C} = 200 MVAr, Q_{9C} = 350 MVAr. All lines has rated voltage 230 kV.

For performing simulations, PSLF in version 17.05 was used, provided to authors by Slovenská elektrizačná a

prenosová sústava (SEPS). At first, power system was modeled in steady state, with values in ohms, or in micro-mhos (B_C values). In next step, dynamic models were added for all generators. First model is model "genrou", it is a model of solid rotor generator represented by equal mutual inductance, its block model is displayed at Fig. 4.

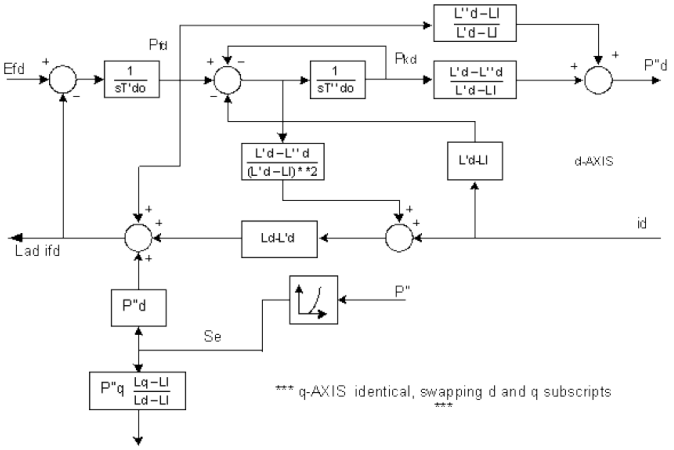


Fig. 4 Block diagram of synchronous machine "genrou" model.

Next model used was model of turbine and governor, "ieeeg1", it's an IEEE steam turbine/governor model. Therefore, machine and prime mover was modeled, and then next two models were added, first was model of exciter, for which was chosen model "esac4a", which is IEEE (1992/2005) type AC4A excitation system model. Block diagram is shown on Fig. 5.

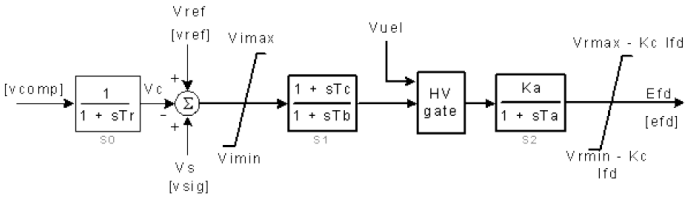


Fig. 5 Block diagram of exciter model "esac4a".

For investigation of PSS behavior, two PSS models were chosen for comparison. First was model "pss1a" which is a single input PSS, and second was model "pss2a" which is a dual input PSS (IEEE type PSS2A). Block models of these PSS are shown on Fig. 6 and Fig. 7.

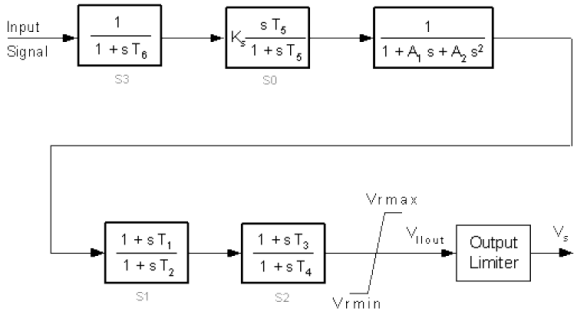


Fig. 6 Block diagram of "pss1a"model

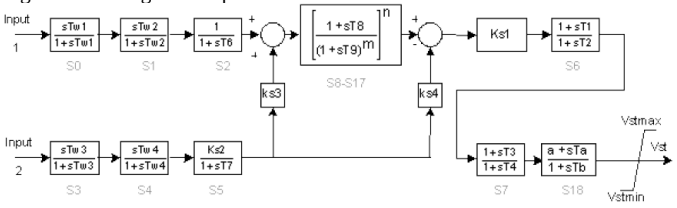


Fig. 7 Block diagram of "pss2a" model

An interconnection between PSS and exciter is made in the accumulator block of exciter through Vs input.

IV. Investigation of PSS influence to stability

Two cases will be presented, in first, an influence of excitation system parameters will be presented, and in second, an influence of two different PSS configuration will be presented. For both cases, there were used same parameters of exciter, which are: T_r = 0,02 s, K_a = 40, V_{iMAX} = 10, V_{iMIN} = -10, T_a = 0,02 s, V_{rMAX} = 10, V_{rMIN} = -10.

In the first case, a "pss1a" model was used, and it was examined an influence of changing stabilizer gain to power system performance during fault. Fault chosen for this examination was three-phase solid short circuit on the one of the lines between buses 8 and 9, very near to the bus 8. Fault occurs at time 1 second, and is cleared after three cycles from both sides, in time 1,06 seconds. For model "pss1a" were same parameters set all the time, except stabilizer gain, which was subject to change. As input signal for stabilizer was used frequency deviation of generator bus voltage. Pss1a parameters enlisted: T_6 = 0, T_5 = 1 s, A_1 = A_2 = 0, T_1 = 0,02 s, T_2 = 0,05 s, T_3 = 5,4 s, T_4 = 3 s, V_{rMAX} = 10, V_{rMIN} = -10. Stabilizer gain was subject of change in the interval from 0 to 40, with step of 5. Not all characteristics were displayed, due to better displaying clarity, only for values of 0 (PSS turned off), 15, 20, 25.

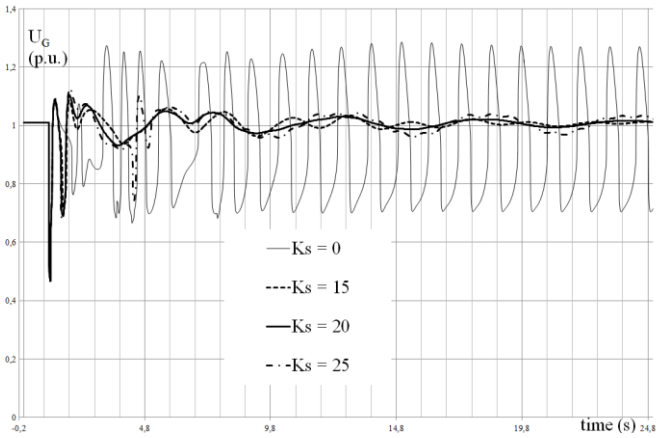


Fig. 8 Variations of generator 4 voltage depending on PSS gain

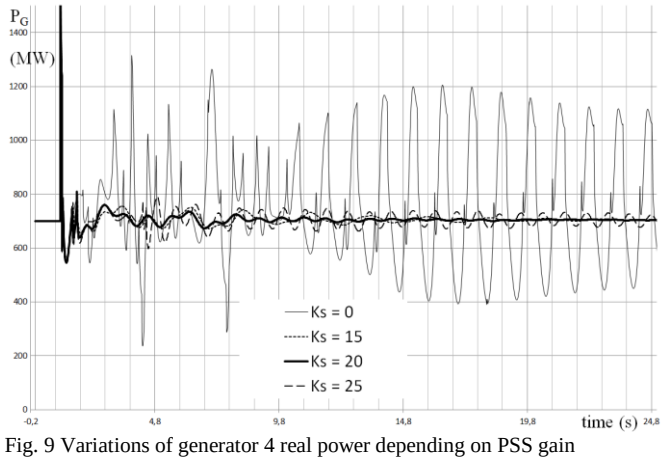


Fig. 9 Variations of generator 4 real power depending on PSS gain

From these figures (from 8 to 10) there can be seen, that the best damping case is observed for K_S = 20. For this particular generator and for this type and place of fault this value can be found as the most convenient. For obtaining right values for any type of fault it is needed to examine worst possible fault in particular power system, and choose setting according that type of fault. [7]

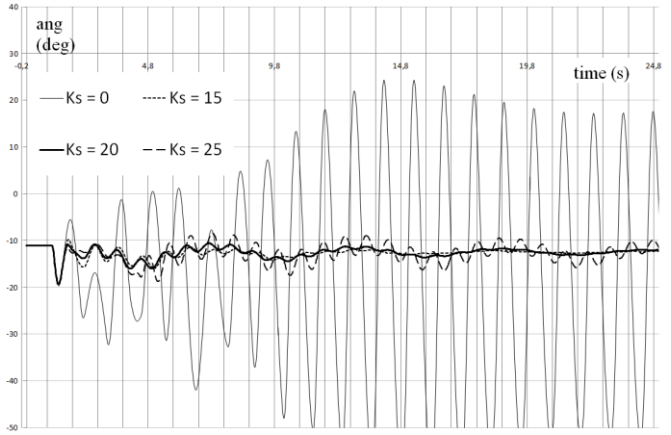


Fig. 10 Variations of generator 4 relative angle (related to generator 3) depending on PSS gain

In the second case, there was two models of PSS used, first was, as in the previous case, model "pss1a", single input stabilizer, the second one was double input stabilizer "pss2a". Used parameters for "pss1a" were exactly the same as in previous case, with optimal K_S parameter value obtained in previous case, therefore $K_S = 20$. Parameters used for "pss2a" model were used as following: $T_{W1} = 10$ s, $T_{W2} = 0$, $T_6 = 0$, $T_{W3} = 2$ s, $T_{W4} = 0$, $K_{S2} = 0$, $T_7 = 0$, $K_{S3} = 1$, $T_8 = 0{,}5$ s, $T_9 = 0{,}1$ s, n = 0, m = 0, $K_{S4} = 0$, $K_{S1} = 2$, $T_1 = 1$ s, $T_2 = 0{,}5$ s, $T_3 = 2$ s, $T_4 = 0{,}1$ s, a = 1, $T_a = T_b = 0$, $V_{stmax} = 0{,}2$, $V_{stmin} = -0{,}2$. Input signal used for input 1 was rotor speed, for input 2 it was generator real power output. Dependence with no PSS was also added. For this examination, same fault was chosen, as in first case.

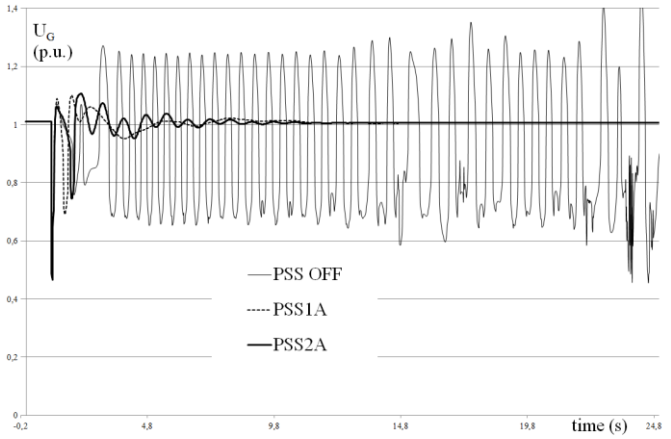


Fig. 11 Variations of generator 4 voltage depending on used PSS model.

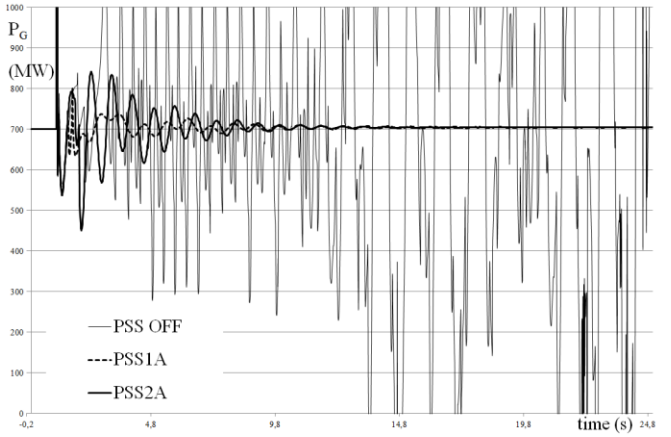


Fig. 12 Variations of generator 4 real power depending on used PSS model.

From these figures there can be seen such a influence of PSS model to oscillation damping. PSS model "pss2a" has more positive influence on oscillation damping than "pss1a", even without really optimal settings, so it is shown, that additional input can be helpful and can improve PSS performance.

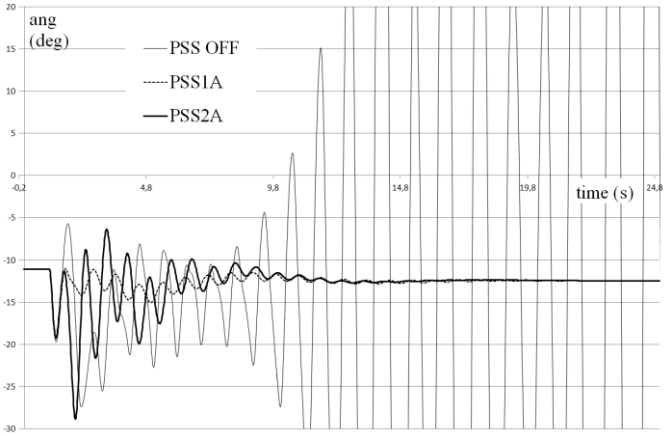


Fig. 13 Variations of generator 4 relative angle (related to generator 3) depending on used PSS model.

V. Conclusion

In this paper, an influence of PSS parameters and PSS model was investigated. The parameters were chosen from defined interval, and optimal parameters for PSS were found. Also, influence of different PSS models was investigated.

Nowadays, there are many different methods, which can be used, for optimizing excitation system parameters, such as genetic algorithms, or dynamic programing. [6],[9] Another possibility is to use real-time tuning of PSS using neuro-fuzzy system, such as IDIPSS, which has been proposed in [8]. There is a very large space for further research of these problems.

Acknowledgment

This work was supported by Slovak Research and Development Agency under the contact No. APVV-0385-07 and VEGA 1/0166/10 and VEGA 1/0388/13 projects.

References

[1] Kundur, P.: Power System Stability and Control. New York: McGraw-Hill, 1998. ISBN 0-07-035958-X.

[2] Kundur, P. - Paserba, J. - Ajjarapu, V. - Andersson, G. - Bose, A. - Canizares, C. - Hatziargyriou, N. - Hill, D. - Stankovic, A. - Taylor, C. - Van Cutsem, T. - Vittal, V.: Definition and classification of power system stability. In: Power Systems, IEEE Transactions, roč.19, č.3, s. 1387- 1401, Aug. 2004. ISSN 0885-8950.

[3] Kolcun, M. - Chladný, V. – Varga, L. – Beňa, Ľ. - Ilenin, S. – Leščinský, P. - Mešter, M.: Analýza elektri-začnej sústavy. Košice. Technická univerzita Košice 2005. ISBN 80-89057-09-8.

[4] Software documentation PSLF 17.0_05. General electric 2010.

[5] IEEE Recommended practice for Excitation System Models for Power System Studies, IEEE Standard 421.5-1992/2005.

[6] CHUVYCHIN, V. – PETRICHENKO, R. – GUROV, N. – DAMBIS, A.: The influence of excitation system parameters to the power system stability. Riga Technical University, Riga, Latvia.

[7] CHUVYCHIN, V. – PETRICHENKO, R. – GUROV, N.: Optimisation of Excitation System Parameters for Kegums Hydro Power Plant of Latvia. Riga Technical University, Riga, Latvia.

[8] SHARMA, A. – KOTHARI, M. L.: Inteligent dual input power system stabilizer. SciVerse, March 2003, Vol. 64, Issue 3, p.p. 257-267.

[9] MURGAŠ, J. – HNÁT, J. – MIKLOVIČOVÁ, E.: PSS Parameters Setting using genetic algorithms. AT&P Journal Plus2, 2008. Pp. 67-70.

Transient stability enhancement using thyristor controlled series compensator

[1]*Zsolt ČONKA (1st year),* [2]*Matúš NOVÁK (2st year)*
Supervisor: [3]*Michal KOLCUN*

[1, 2, 3] Dept. of Electric Power Engineering, FEI TU of Košice, Slovak Republic

[1]zsolt.conka@tuke.sk, [2]matus.novak@tuke.sk, [3]michal.kolcun@tuke.sk

***Abstract*— This paper describes the improvement of transient stability limit of power systems with the utilized TCSC. Improving the transient stability of electricity transmission is one of the main tasks of these devices. Term improvement of transient stability means the increase maximum transmission capacity without loss of synchronization. Ability to improve the transient stability by TCSC is shown on a simple power system.**

***Keywords*—Power system stability, TCSC, transient stability.**

I. INTRODUCTION

One of the most common controllers used to compensate of damping oscillations in a power system is a power system stabilizer (PSS). Nowadays we can use the additional devices for these purposes. The best devices for improving transient stability are FACTS devices. These devices were designed for different purposes. One of the most popular FACTS devices is Thyristor-controlled series compensator (TCSC). Main use of this device is to control the power flow in the transmission system. But there are other benefits that can be exploited to improve the transient stability of the transmission system [1].
The Unified Power Flow Controller is one of the promising devices in the FACTS family since it can control three basic parameters of power flow for AC transmission namely voltage magnitude, phase angle and line impedance. A number of FACTS controllers based on the rapid development of power electronics technology have been proposed in recent years for better utilization of existing transmission facilities [1].

II. THYRISTOR CONTROLED SERIES COMPENZATOR (TCSC)

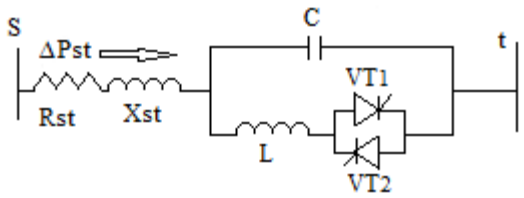


Fig.1 TCSC [3]

The Fig.1 shows the main diagram of TCSC. The switching elements of the thyristor controlled reactor (TCR) consist of two anti-paralleled thyristors, which alternate their switching at the supply frequency [2]. The system is controlled by varying the phase delay of the thyristor firing pulses relative to the zero crossings of some reference waveform. The effect of such variation can be interpreted as a variation in the value of the capacitive or inductive reactance at the fundamental frequency. In our analysis, the thyristor will be ideal, so that nonlinearities due to the thyristor turn on and turn-off is neglected. We also assume that the limit current $\boldsymbol{i_e}$ is essentially sinusoidal, and take it to be the reference waveform for the synchronization of the firing pulses a slight modification of our analysis is needed if the synchronization $\mathbf{i_s}$ instead done with the capacitor voltage V [2].

This model utilizes the concept of a variable series reactance which is adjusted through appropriate variation of the firing angle (α). The controller comprises of a gain block, a signal washout block and a phase compensator block. The input signal is the normalized speed deviation (Δν), and output signal is the stabilizing signal (i.e. deviation in conduction angle,Δσ) [3].

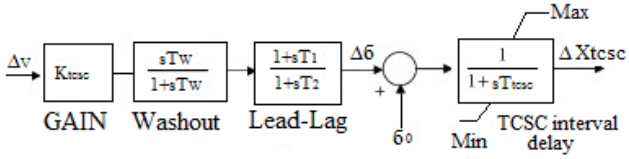


Fig. 2 TCSC control [3]

TCSC controller model can be represented by the following state equations; [3]

$$\Delta\dot{\alpha} = -\frac{1}{T_2}\Delta\alpha - \frac{K_{tcsc}}{\omega_s}\left(\frac{1}{T_2}\right)\Delta\omega - \frac{K_{tcsc}}{\omega_s}\left(\frac{T_1}{T_2}\right)\Delta\dot{\omega} \quad (1)$$

$$\Delta\dot{X}_{tcsc} = -\frac{1}{T_{tcsc}}\Delta\alpha - \frac{1}{T_{tcsc}}\Delta X_{tcsc} \quad (2)$$

The steady-state relationship between the firing angle α and the equivalent TCSC reactance, Xtcsc is described by the following relationship is [3]

$$X_{tcsc} = -X_C + C_1(2(\pi-\alpha) + \sin(2(\pi-\alpha))) - \\ - C_2\cos^2(\pi-\alpha)(\omega\tan(\omega(\pi-\alpha)) - \tan(\pi-\alpha)) \quad (3)$$

where

$$X_{LC} = \frac{X_C X_L}{X_C - X_L} \quad (4)$$

$$C_1 = \frac{X_C + X_{LC}}{\pi} \tag{5}$$

$$C_2 = \frac{4X_{LC}^2}{\pi X_L} \tag{6}$$

The linearized TCSC equivalent reactance [3]

$$\Delta X_{tcsc} = -2C_1(1+\cos(2\alpha)) + C_2\sin(2\alpha)(\varpi\tan(\varpi(\pi-\alpha)) - \tan\alpha) + C_2\left(\varpi^2\frac{\cos^2(\pi-\alpha)}{\cos^2(\varpi(\pi-\alpha))} - 1\right)\Delta\alpha \tag{7}$$

III. TRANSIENT STABILITY

Synchronous machines are mainly used in the PS as synchronous generators, synchronous motors and synchronous compensators. These devices are interconnected through transformers and lines and they are in parallel and synchronous operation. The transfer capability of such systems is limited by the permissible voltage drops and with the power handling capacity. At transmission to the large distance been treated the condition of stability of parallel operation. Prerequisite for the existence and operation of large interconnected power systems is the steady stability of parallel operation that is to say sync. Synchronization power of synchronous machine allows us to so synchronous operation. This power presents the increase of the transmitted power at an increasing of the load angle of generator rotor over 1 ° [4], [5]. The swing of generator rotor due to changes of the electromagnetic energy which is accumulated in the magnetic circuits of the machine so arises additional power. This power can take positive but also negative values, which is reflected as a contribution to the performance of ΔP as a braking, or acceleration of the machine. [4], [5].

The role of synchronous machines is not the retention of the synchronism. Co-operation of synchronous machines cannot be at arbitrarily large power. The Steady-state operation of whole system is depends on the electrical, mechanical and electromagnetic parameters of the system [4], [5].

If you are not violated conditions of transfer we can reach a stable maximum output with the smoothly increasing transmitted power. The considerations of static stability we can apply for small swings of the machine. In operation, there are many sudden changes in the system (switching processes, shock loads, short circuits). Due to the unbalance of the consumption and production of the electric power the load angle values can achieve major changes. These marches change the load angle with leap. After these changes the system goes into the new system state with electromechanical oscillations. Inertia of the machine not allow an immediate change of operating parameters (angle δ) course of these oscillations may be such as that the angle stabilizes at a new constant, or continue to increasing. In the latter case there is a loss of stability. In examining of the transient stability is generally assumed a constant value of the transition electromotive alternator voltage E [4], [5].

IV. SIMULATION

The impact of TCSC on the transient stability is shown on 4 machine system.

At time t = 0.1 s will due to short-circuit failure of one of the parallel lines. Due to the failure comes to a rotor swings. In the first case, the simulation is without TCSC and in the second case involved the TCSC.

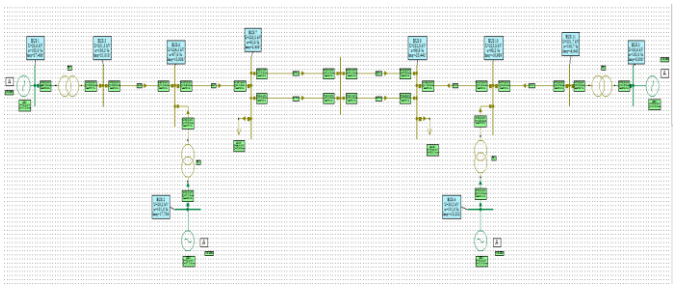


Fig 3. 4 machine system without TCSC

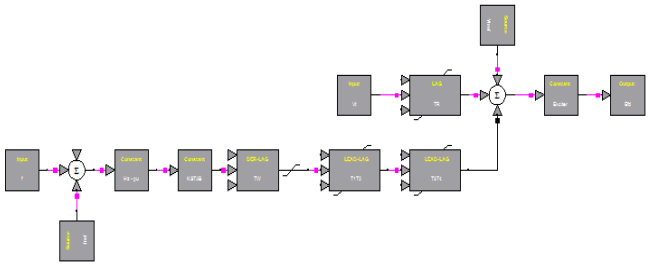


Fig. 4 Automatic voltage regulator and power system stabilizer

Each generator has AVR (Automatic voltage regulator) and Power system stabilizer (PSS). Fig. 4 shows the regulator which contains PSS and AVR too.

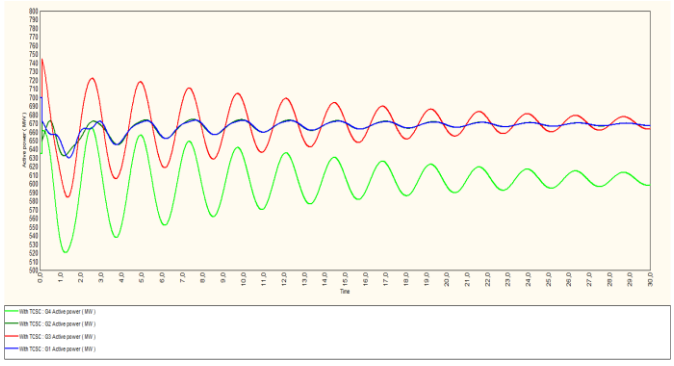


Fig.5 Real power (MW) swings of the generators without TCSC

Fig. 5 describe that the real power (MW) of the generators become to swings and the settling time of the fluctuations are very long. The big generators rotor swings and the long settling time adversely affect the stability of the power system, because another failure could lead to the fault of one or more generators.

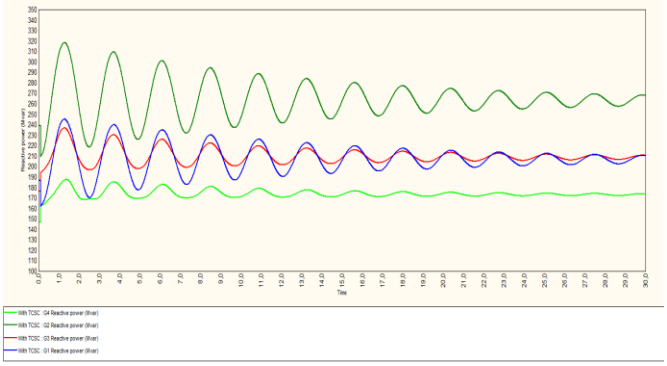


Fig.6 Reactive power (MVar) swings of the generators without TCSC

The swing of the reactive power of the generators impacts on the voltage of the system. Long and big swing of reactive power leads to a voltage instability.

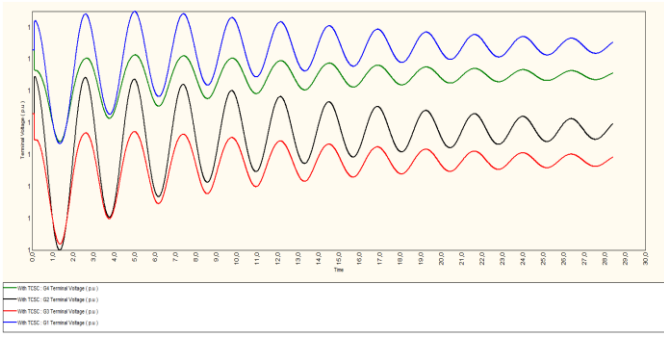


Fig.7 Terminal voltage (p.u.) of the generators without TCSC

Fig. 7 shows the voltage fluctuations on the generators terminal.

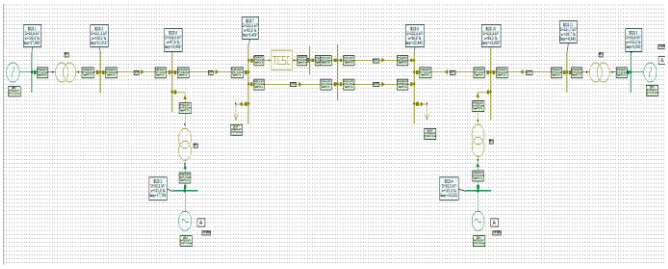


Fig 8. 4 machine system with TCSC

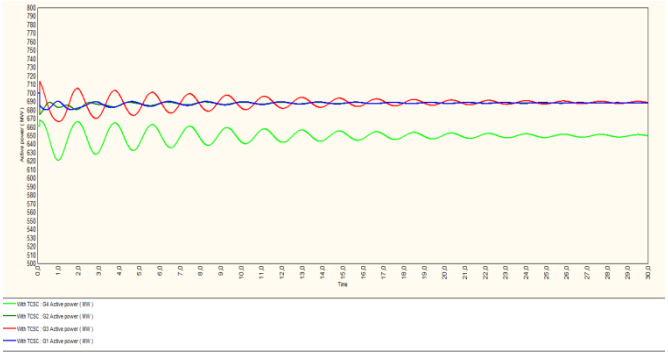


Fig. 9 Real power (MW) swings of the generators with TCSC

After the same fault as without TCSC, the real power of the generators swing less than without TCSC. Generators have a shorter settling time than without TCSC.

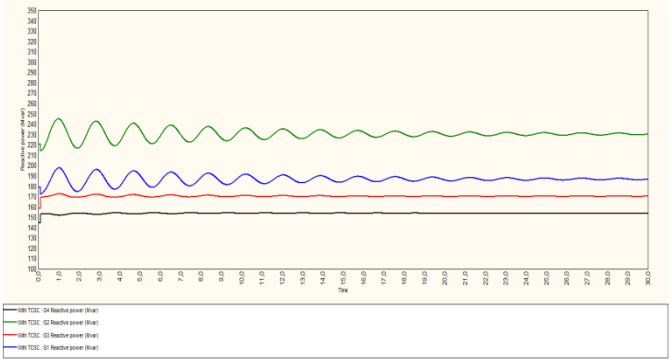


Fig. 10 Reactive power (MVar) swings of the generators with TCSC

Reactive power swings and settling time is reduced by using TCSC.

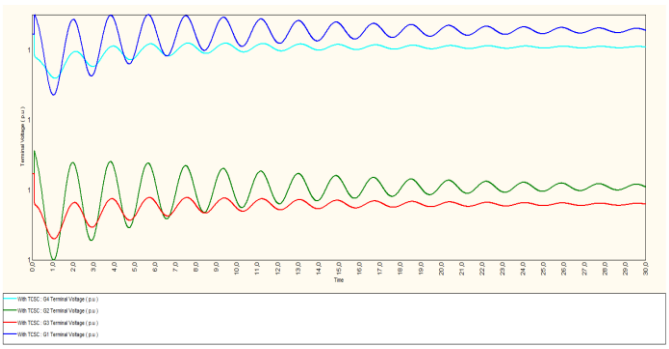


Fig.11 Terminal voltage (p.u.) of the generators with TCSC

V. Conclusion

Nowadays, when the number of non-predicable renewable energy sources constantly grown, we need to increase the transient stability of power transmission. One way for reach better stability is utilizing of FACTS devices. This model shows the positive effect of the TCSC on the transient stability of the power system and can help to get the better reliability of electricity supply.

Acknowledgment

This work was supported by Slovak Research and Development Agency under the contact No. APVV-0385-07 and VEGA 1/0166/10 and VEGA 1/0388/13 projects.

References

[1] N.M. Jothi Swaroopan, P. Somasundaram, Transient stability enhancement with UPFC http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4735849&tag=1W

[2] P. Mattavelli, G.C. Verghese, A.Stankovic, *Phasor Dynamics of Thyristor-Controlled Series Capacitor Systems* [online]. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1459995

[3] D. Mondal, A. Chakrabarti, A. Sengupta, *Optimal placement and parameter setting of SVC and TCSC using PSO to mitigate small signal stability* problem [online]. http://www.sciencedirect.com/science/article/pii/S0142061512001123

[4] TROJÁNEK, Zdeněk – HÁJEK, Josef – KVASNICA, Pavol: Přechodné jevy v elektrizačních soustavách. STNL , 1987. 202 – 231 s.

[5] REISS, L., MALÝ, K., PAVLÍĆEK, Z., BÍZIK, J.: Teoretická elektroenergetika II. STNL Praha, Alfa Bratislava 1978

[6] M. A. ABIDO. Power system stability enhacement using FACTS controllers [online]. [cit. 2012-02-05]. Available on the Internet: <http://ajse.kfupm.edu.sa/articles/341B_P.12.pdf>

[7] FACTS – Flexible AC Transmission Systems: Series Compensation. Erlangen: Siemens AG, 2010. [online]. < http://www.energy.siemens.com/us/pool/hq/powertransmission/FACTS/FACTS_Series_Compensation.pdf>.

[8] FACTS – powerful systems for flexible power transmission [online]. <http://www05.abb.com/global/scot/scot221.nsf/veritydisplay/b0f2c8c94b48a6bcc1256fda003b4d42/$file/facts%20eng.%20abb% 20review.pdf >

[9] NARAIN, G. HINGORANI, LASZLO, GYUGYI,: Understanding FACTS. ISBN: 0-7803-3455-8, IEEE Order No. PC5713

Usage of Optical Correlator in Traffic Sign Inventory System

[1]*Tomáš HARASTHY (3rd year)*
Supervisor: [2] *Ján TURÁN*

[1,2]Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

[1]tomas.harasthy@tuke.sk, [2]jan.turan@tuke.sk

***Abstract* — This paper describes Traffic Sign Inventory System based on Optical Correlator. The tested video has been captured by color camera and then it's controlled by proposed system. Proposed system contains following main blocks: Capturing video, Correlation process and Identification. First block consist of two main parts: color capturing and preprocessing of captured video. Correlation process is done in the next block. This block consists of producing input plane of correlation process and correlation plane. The third block consists of identification block which follows correlation plane, concretely Region of Interest of whole correlation plane. Block and hardware scheme have been proposed and they are specified below.**

***Keywords*— Traffic Sign Inventory System, Optical Correlation, Region of Interest**

I. Introduction

Nowadays are Traffic Signs Recognition Systems very popular and often cited topics in a field of info-electronics. There are many ways and systems to recognize traffic signs which can compare video stream with database of traffic signs and recognize capture traffic signs. This paper is not concentrated to Traffic Signs Recognition System, which was described in other manuscripts, but it is focused to great using of this system. Traffic Signs Recognitions Systems are most focused to secure driver and crew of car, but there are more fields where could be these systems used. In this paper is designed and described to inventory traffic signs on chosen road [1,2,6,7,8].

Video is captured with color camera. Captured video is then preprocessed (color filtered). Video is continuously recorded with GPS and these coordinates are assigned to each Traffic Sign. GPS coordinates give information about position of Traffic Signs, where the searched traffic sign should be situated. Proposed system is interconnected with database of Traffic Signs for chosen road. Database contains all Traffic Signs which supposed to be on the road and according to GPS coordinates are traffic sign chosen and searched in captured video. Traffic Sign from captured video is using by Optical Correlator to compare with Traffic Sign from database and if match is occurred, Traffic Sign from road is marked as readable [1,3,4,]. If match does not occur, Traffic Sign, which supposed to be on monitored place is damage, or missed.

In Chapter 2 is design of whole system described. Main parts of scheme are detailed described. Chapter 3 contains experiments with proposed system. Conclusion and results are covered in Chapter 4.

II. Design of Traffic Sign Inventory System based on Optical Correlator

A. Hardware Design

Traffic Sign Inventory System based Optical Correlator has 5 main parts: Color Camera, Computer, GPS module (can be a part of camera), Optical Correlator and Display. In Fig.1 the hardware scheme of proposed system is shown.

In first, video is captured by the color camera. Video camera is situated in the car and has to be focused to right part of road. Camera in this position will record the place where the occurrence of traffic signs is the most common.

Continuously with video from camera are GPS coordinates recorded as well. GPS coordinates give us information where the vehicle is currently located and that gives information which reference traffic sign from database is needed [5].

In database are recorded all Traffic Signs which should be located on chosen road. Database contains expect reference Traffic Signs GPS coordinates of each Traffic Signs and their detailed description.

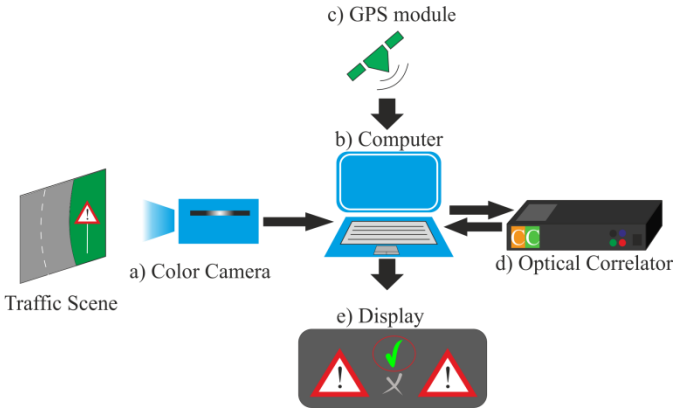


Fig. 1. Hardware scheme of Traffic Sign Inventory System based Optical Correlator.

Captured Traffic Scene from digital camera is preprocessed in Computer with color filter and Region of the Interest (ROI) is defined. ROI contains a potential Traffic Sign. This chosen ROI is then compared using Optical Correlator.

If match occurs, monitored Traffic Sign and reference Traffic Sign are the same and System take another

Traffic Sign from database, according to GPS coordinates.

If match between reference and captured (ROI) Traffic Signs does not occur, captured Traffic Signs is damaged or missing and system marks traffic sign on this position as Invalid.

B. Block Scheme

Hardware scheme described above contains only hardware components which are necessary to build our system. Block and functional scheme is shown in Fig. 2.

Video is captured with color camera and then is preprocessed in computer. First step of preprocessing is correction of exposure. In next step of preprocessing, the color filter is used. Color filters are designed to remove all irrelevant colors in input traffic scene. Input image is in HSV color space.

The color segmentation is processed. Every traffic sign has his dominant color. On the Slovak roads most often Yellow, Red and Blue color are used. This means we need to create three binary maps, one for each of these colors [4]. By analyzing hue component (H), we can identify Blue, Yellow and Red regions in our processed image.

For each image pixel is possible to use hue-based detection for Blue, Red and Yellow color. For each color following equation are used

$$Y = e^{\frac{-(x-42)^2}{30^2}} \quad (1)$$

$$R = e^{\frac{-x^2}{20^2}} + e^{\frac{-(x-255)^2}{20^2}} \quad (2)$$

$$B = e^{\frac{-(x-170)^2}{30^2}}. \quad (3)$$

Equations Y gives values close to 1 for Yellow regions, R gives values close to 1 for Red regions and B gives values close to 1 for Blue regions. In this equations we can see, that H can be from range 0-255. Yellow can be detected near value 42, Red near values 0 and 255 and Blue value is 170.

These equations can be tuned for every other needed color. After threshold for detected colors three binary maps (Red, Blue and Yellow) are created. Color filtering removes all irrelevant colors. Color filtering get image with color clusters, which are irrelevant for detecting Traffic Sign. In next step in preprocessing, irrelevant color clusters are removed. After color filtering and irrelevant cluster removing, the potential Region of Interest (ROI) is extracted. The potential ROI is defined at the input image [4,5].

Continuously from database is chosen Traffic Sign which is expected according to GPS coordinates. When is vehicle moving, GPS coordinates determines which Traffic Sign from database is expected. From reference Traffic Sign, (from database) and ROI from processed video is input scene for optical correlation produced.

Input Scene is input plane for correlation process and correlation is done in Optical Correlator. In our design is Joint Transform Correlator used. This type of correlator does not need reference filter, (in face to Matched Filter Correlator), so correlation is done in real time. Input scene of correlation process is done from ROI from captured video and Traffic Sign from database.

Output of correlation process consists of pair of correlation peaks per match. Pair of correlation peaks represent match in input plane, so Traffic Sign from database and captured Traffic Sign from road are same. After this correlation enter to input plane of correlation process next pair of traffic signs, next sign from database and next sign from captured video. If in correlation plane do not occur cross-correlation peaks, Traffic Signs in input plane are not same [10].

Result of this correlation is, that captured scene does not contain correct Traffic Sign, so Traffic Sign can be damaged or missed. This traffic sign (from database) is marked and needs to be checked.

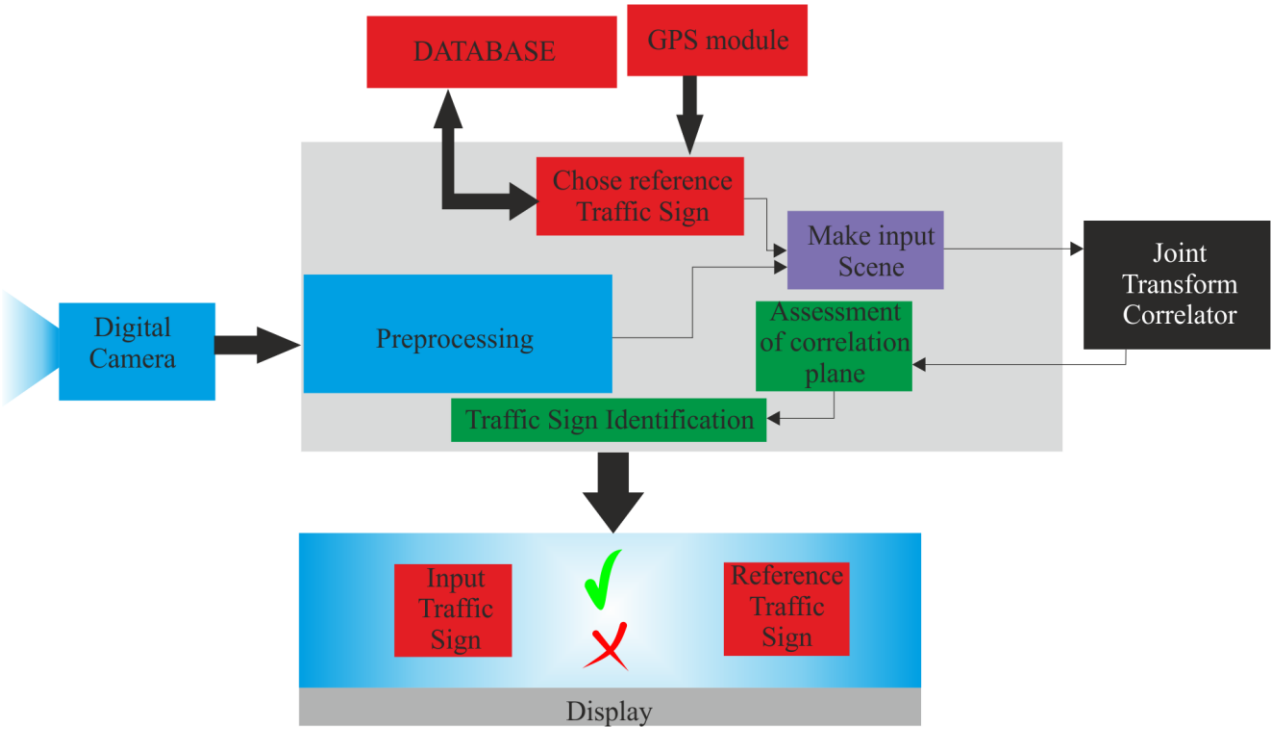


Fig. 2. Block scheme of Traffic Sign Inventory System based on Optical Correlator.

III. EXPERIMENTS

Experiments with this system were done on Highway D1, between Košice and Prešov. There have been captured video and some Traffic Signs from this road are shown in Fig. 3.

Fig. 3. Captured Traffic Signs.

Chosen Traffic Signs have been compared using Joint Transform Correlator with origin Traffic Signs from database which is showed in Fig. 4.

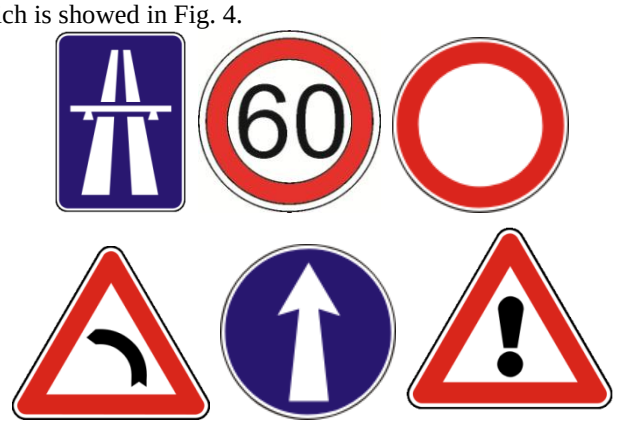


Fig. 4. Traffic Signs from database.

Database includes Traffic Signs shown in Fig. 4, and information about each traffic sign. Captured Traffic Signs are shown in Fig. 3. Order of shown traffic signs is correct, so comparing (correlation) of this signs should give information, that all tested traffic signs are OK and no traffic sign is damaged or missed. Results of this system are recorded in TABLE I. In this table are recorded intensities of correlation peaks (two peaks for match). Threshold of success was set to value 180, so if average intensity of correlation peaks was higher than threshold value (180), tested traffic sign was marked as OK. If average intensity of correlation peaks was lower then 180, traffic sign should be marked as "FAULT", so traffic sign in video is damaged or missed

TABLE I
RESULTS OF CORRELATION

TRAFFIC SIGN	Intensity of 1. peak	Intensity of 2. peak	Average of Intensities (255 – 100%)	Result
Start of Highway	204	202	79%	OK
Speed Limit (60)	220	225	87%	OK
Do not enter	184	182	71%	OK
Bend to Left	243	242	95%	OK
Ahead only	222	220	86%	OK
Other danger	194	191	75%	OK

IV. CONCLUSION

In this paper was Traffic Sign Inventory System based on Optical Correlator described. Proposed system uses recorded video to control Traffic Signs on highway D1 in Slovakia between Košice and Prešov. Video was captured with color camera and preprocessed in computer. Preprocessing of video consists of ROI. ROI contains possible Traffic Signs and these were compared with database.

Database contains information about all traffic signs situated on mentioned road. In this paper is tested one part of highway. Experiments were done with six traffic signs. Traffic Signs were chosen from database according to GPS coordinates from GPS module. To make system more compact we can use video camera coupled with GPS module.

In tested video were recognized and controlled all tested Traffic Signs. If system marks some traffic sign as unrecognized, Traffic Sign is mark as a damaged or missed. These Traffic Signs need to be check and replaced or repaired.

In our case all Traffic Signs was marked as OK and no reparation is needed.

REFERENCES

[1] J. W. Goodman, "Introduction to Fourier Optics," Colorado: Roberts & Company, Publishers, 2005.

[2] Yu, F. T. S. and Yang, X.: "Introduction to Optical Engineering," Cambridge University Press, 1997.

[3] F.P. Paulo and L.P. Correia, "Automatic detection and classification of traffic signs," Image Analysis for Multimedia Interactive Services, WIAMIS, 6-8 June 2007.

[4] S. E. Umbauch, "Computer Vision and Image Processing and A Practical Approach Using CVIPtools", Prentice Hall PTR, Upper Saddle River, NJ 07458, June 1999.

[5] T. Harasthy, J. Turan, L. Ovsenik and K. Fazekas, "Traffic Signs Recognition with Using Optical Correlator", Proc. IWSSIP 2011, Sarajevo, Bosnia and Herzegovina,16-18 June 2011, 239-242, 2011.

[6] D. S. Kang, N. C. Griswold and N. Kehtarnavaz, "An Invariant Sign Recognition System Based on Sequential Color Processing and Geometrical Transformation," Department of Electrical Engineering, Texas A&M University, College Station, Texas, 1994.

[7] P. Ambs, "Optical Computing: A 60-Year Adventure, " Hindawi Publishing Corporation Advances in Optical Technologies, Volume 2010.

[8] L. Guibert, Y. Petillot and J.-L. de Bougrenet, "Real-time demonstration of an on-board nonlinear joint transform correlator system", Society of Photo-Optical Instrumentations Engineers, Optical Engineering, Vol. 36 No. 3, Brest Cedex, France, March 1997.

[9] Cambridge Correlators, http://www.cambridgecorrelators.com, December 2012

[10] L. Guibert, G.Keryer, A. Servel, M. Attia, H. S. MacKenzie, P. Pellat-Finet, J.-L. de Bougrenet de la Tocnaye, "On-board optical joint transform correlator for real-time road sign recognition", Optical Engineering, Vol. 34 No 1, France, January 1995.

Use of the Robotics Toolbox in the creation of GUI in MATLAB

[1]*Marek VACEK (2st year – PhD.),* [2]*Martin LEŠO (1st year – Ing.)*
Supervisor: [3]*Jaroslava ŽILKOVÁ*

[1]Dept. of Electrical Engineering and Mechatronics, FEI TU of Košice, Slovak Republic

[1]marek.vacek@tuke.sk, [2]martin.leso@student.tuke.sk, [3]jaroslava.zilkova@tuke.sk

***Abstract*—The article describes the process and possibilities of creating a graphical user interface for use in robotics, using a freely downloadable toolbox for MATLAB – the Robotics Toolbox. This Toolbox is used to calculate forward and inverse kinematics and to create a simulation model of a robotic arm. The final graphical user interface can be used for research or for teaching purposes.**

***Keywords*—Robot manipulator, Robotic arm, Robotics Toolbox, Graphical User Interface.**

I. INTRODUCTION

The following article describes the process of creating a graphical user interface (GUI) to control a model of robotic manipulator in MATLAB. Robotics Toolbox, which is freely downloadable and offers a wide variety of functions for calculations in robotics, will be used to do this. As Robotics Toolbox can only be controlled via the command line in MATLAB, it is useful to create a user interface to directly enter the position data of the given robotic arm model. Therefore, we created a GUI in which the entering of the data and the presentation of the results are much quicker and more user-friendly. The final calculated values from the GUI will be used in our further endeavors to calculate and generate control packets for robotic arm servomotors. Other functions will be added to the final GUI. The GUI will control and monitor an optical security system in a robotic workstation.

II. GUI DESIGN

Before the creation of the GUI itself, it is necessary to stipulate requirements for the interface. We stipulated them as follows:

- rendering of a 3D robotic manipulator model for the given position of joints
- solving forward kinematics by entering the angles of individual robot joints or by changing the position of the sliders
- displaying the transformation matrix required and the current transformation matrix of the effector.

Based on these requirements, it is beneficial to draw the layout of the GUI elements before the actual creation of the GUI (Fig.1).

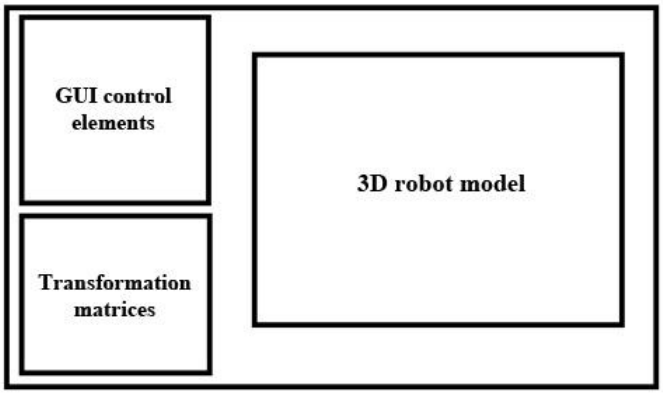


Fig. 1. Layout of GUI control elements

III. GUI CREATION

The GUIDE editor was used to create the GUI. It can be run in MATLAB by entering the command "guide".

The elements were then placed into the empty window in accordance with the design. To achieve a clear arrangement of the GUI elements and to be able to modify the GUI faster, those elements which have a similar function should be placed together into one unit.

- PushButton - button performing an action after being pushed
- Edit text - text typing
- Axes - displaying of graphs and pictures
- Slider - numerical input for a certain range
- Static Text - displaying of static text
- Table – displaying of tables
- Panel - panel into which objects are inserted

In the editor, the placement of elements according to our design looks like this (Fig.2).

IV. SETTING OF THE GUI ELEMENTS PROPERTIES

"Property Instructor" was used to show the properties of individual elements. It can be accessed by double clicking the object required. "Tag" is one of the basic properties of objects and it defines the name of the element. It is necessary for all the elements which perform some action to be properly named because when programming the GUI, the names of these elements are used in the program. The properties of those

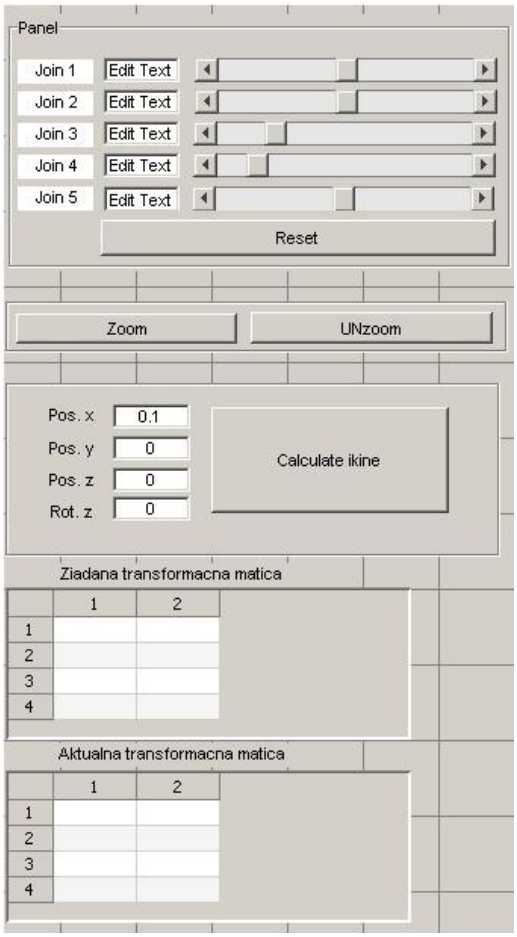


Fig. 2. GUI implementation – closer view of control elements

elements which are important from the point of view of functionality and GUI display follow.

The "Static Text" element displays static text, and it is, therefore, necessary to set the String parameter for the required displayed text of an element.

For the "Slider" element, it is necessary to set the properties of "Min" and "Max", which give us the minimum/maximum value of the slider.

"Pushbutton" must be set to the "String" parameter, which serves to display the name of the button.

The last element for which parameters must be set is the "Table" element. There is a "ColumnName" option in the "Property Inspector" window. This option opens a new window called "Table property editor", in which parameters such as the number, size, and name of individual rows and columns can be set.

V. GUI PROGRAMMING

After all properties have been set, it is necessary to save the GUI under a name of your choice. After its saving, an .m file, into which all the functions of individual objects are programmed, is automatically generated.

Robotics Toolbox was used to create the GUI. Robotics Toolbox is a toolbox for MATLAB. It offers a lot of functions which are useful in the field of robotics, including kinematics, dynamics, and trajectory calculation. It can also be used for simulations, as well as for analysis of the results of experiments with real robots.

From among all the functions, the following were predominantly used:

Ffkine - forward kinematics calculation
Ikine - inverse kinematics calculation
SerialLink – serial robotic manipulator creation

VI. SERIAL ROBOT MANIPULATOR MODEL CREATION

The robotic manipulator is defined by the function robot=seriallink (Robotics Toolbox function). This function creates the robotic arm object which can be simulated. The DH variable is a matrix in which Denavit-Hartenberg (DH) parameters are defined for n joints as follows (1):

$$DH = \begin{bmatrix} \theta_1 & d_1 & a_1 & \alpha_1 \\ \theta_2 & d_2 & a_2 & \alpha_2 \\ . & . & . & . \\ . & . & . & . \\ \theta_{n-1} & d_{n-1} & a_{n-1} & \alpha_{n-1} \\ \theta_n & d_n & a_n & \alpha_n \end{bmatrix} \quad (1)$$

The individual variables in the matrix are determined via DH notation, which is defined as follows:

- Connection angle: angle θ between the x_{n-1} and x_n axes around the z_{n-1} axis (the rotary joint variable)
- Robot link distance: d – the distance between the x_{n-1} and x_n axes along the z_{n-1} axis (the gliding joint variable)
- Link length: a – the distance between the z_{n-1} and z_n axes along the x_n axis (a constant)
- Link rotation: angle α between the z_{n-1} and z_n axes around the x_n axis (a constant)

A serial robotic manipulator with n links comprises n+1 components. The numbering of components starts at zero, which represents the non-mobile component of the robot, and continues to n, i.e. the last component of the robot. The numbering of links starts at 1, i.e. the link between the first mobile component and the non-mobile component of the robot, and continues to n. Component (i) is linked to component (i-1), and this represents the link number i.

Three conditions must be met when creating the robot:

- All links are defined by the z axis. The z axis is aligned with the connection axis of the joint. The z_n axis must have a positive direction. Determination of the z axis for the rotary joints is therefore obvious. In case of gliding joints it is possible to choose an axis parallel to the direction of gliding.
- The x_n axis is defined along the normal towards the z_{n-1} and z_n axes.

- The y_n axis is defined by the right-hand rule $y_n = z_n \times x_n$

Next, it is necessary to define two variables, for example:

```
robot    % structure including the 3D robot model
RT_robot % model created in Robotics Toolbox
```

The "robot" structure contains a 3D robot model and other data, such as the current values of rotation of individual joints, the transformation matrix of the robot's effector, and the links to the patch objects rendered. We then save these two variables by typing the command "save".

When we run the GUI, the first function to run is the Name_OpeningFcn. This function contains commands which must be carried out when the GUI is started. In our case, the global variables robot and RT robot were defined first. Then, using the "load" command, the variables were loaded, and, using the "plotscene" function, the manipulator was rendered into the "axes" type object.

As concerns the GUI functionality, the Object_Callback type functions are important. These functions are used if, for example, a button is pushed or a slider is used. These actions cause changes in the GUI or change the individual properties of objects. The values of the functions' properties can be acquired by the command get(handles.Tagobjektu,'Object properties') or by setting the property of the function by set(handles.Tagobjektu,'Object properties',value). The most frequent ways of using the Callback functions follow.

Using the Callback function of a button can be illustrated by the following example. We need to calculate the inverse kinematics of a robot. By pushing the button "Calculate ikine", the Callback function of the button will start, in which the following code will be performed:

```
T=loader() % loading the required transformation
matrix
q=iikine(T)  % calculation of inverse kinematics
ffkine(q)  % calculation of forward kinematics and
the 3D model
plotscene()  % 3D model rendering in Axes
rewriter(q)  % writing the calculated values into
the GUI
```

It is also possible to use one's own user functions. If these functions are to load the required data from the GUI, it is necessary to define the handles variable of the function as follows:

```
handles=guidata(gcf());
```

The Callback "edit text" function is used in the following way. After entering the values of rotation of the third joint into the corresponding element and after pushing the Enter button, the Callback function was initialized. In this case, the value of the element was loaded using the following commands:

```
global robot
robot.q(3)=str2num(get(handles.Tagobject,'String')
);
ffkine(robot.q);
plotscene();
```

We can see that when loading the value of the variable using the get command, a text chain entered in the edit text was loaded. This text chain was then converted into a numerical form via the str2num function.

The last example of using the Callback function is the use of the slider. A certain value for the connection angle was set using the slider. In this case, it was possible to proceed in the same way as in edit text, but the difference is that the value of the connection angle was loaded as follows:

```
robot.q(3)=get(handles.Tagobject,'Value')
```

VII. Conclusion

This article described the procedure of graphical user interface creation using the functions of a MATLAB toolbox – Robotics Toolbox. The toolbox was used to create a robotic arm and to calculate the forward and inverse kinematics. The graphical user interface created enables the user to directly enter and calculate the data about the effector's position. The GUI will be modified and adjusted, and used in the future to control an experimental robotic arm. In its current form, it can be used as a teaching resource in robotics classes. To conclude, we enclose a screenshots of the final GUI (Fig.3).

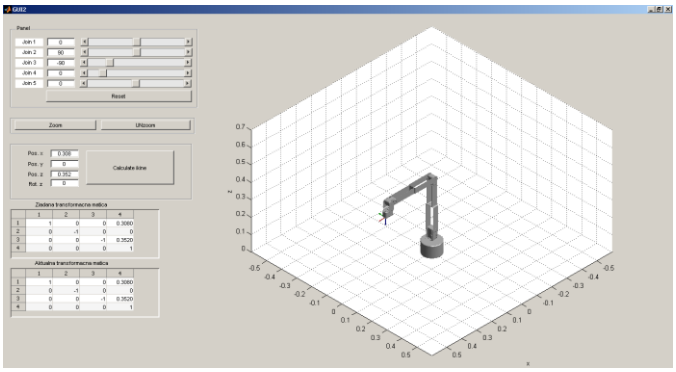


Fig. 3. Final GUI

Acknowledgment

The work has been supported by project KEGA 011TUKE-4/2013.

References

[1] VACEK,M.: Modelovanie priemyselnách robotov, Technical university of Košice, Faculty of Electrical Engineering and Informatics, Diploma Thesis, 2011, (In Slovak)
[2] VACEK,M.: Optický bezpečnostný system robotického pracoviska, Technical university of Košice, Faculty of Electrical Engineering and Informatics, Postgraduate Thesis, 2013
[3] LEŠO,M.: Vizualizácia robotov, Technical university of Košice, Faculty of Electrical Engineering and Informatics, Bachelor Thesis, 2012, (In Slovak)
[4] Mathworks MATLAB: R2012a Documentation [online]. Available on the Internet: <http://www.mathworks.com/help/techdoc/>.
[5] Corke, Peter: Robotics Toolbox for MATLAB, Release 9. Available on the Internet: <http://www.petercorke.com/RTB/robot.pdf>.
[6] Corke, Peter: Robotics, Vision and Control. Springer-Verlag Berlin Heidelberg 2011. 570 s. ISBN 978-3-642-20143-1.
[7] Reza N. Jazar: Theory of Applied Robotics. Springer New York Dordrecht Heidelberg London 2007. 893 s. ISBN 978-1-4419-1749-2.

Using RF system as backup link for FSO systems

[1]*Matúš TATARKO (2nd year)*
Supervisor: [2]Ľuboš OVSENÍK

[1,2]Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

[1]matus.tatarko@tuke.sk, [2]lubos.ovsenik@tuke.sk

***Abstract*—This paper gives us information about hybrid FSO/RF link, where radiofrequency link is used as backup link for primary Free Space Optics link. FSO is optical wireless technology which uses optical laser beam for communication and transmission of data. This type of communication link offers gigabit data rates and low system complexity. Transmission medium of laser beam is air, so atmospheric conditions have negative effects on transmitted signal. Conditions which cause bad influences on FSO signal are negligible for RF signal and reverse. Both links are complementary. In hybrid FSO/RF links is able to combine advantages of both systems and achieve higher overall link availability.**

***Keywords*—Availability, fog sensor, FSO system, RF system.**

I. Introduction

Imagine a technology which offers full duplex connectivity, which can be installed license free worldwide and can be installed in less than one day. In addition it is based on infrared optical, for human invisible, beams of light to provide optical bandwidth connections. That technology is called free space optics (FSO). This technology requires line of sight (LOS) between two static transceivers. Data rates can be from hundred of megabits to tens of gigabits [1,2].

Hybrid FSO/RF communication link contains three basic parts. First part consists of FSO link, second part is RF link and third part is switch or router providing switching between FSO and RF system. FSO is a primary link and RF link is a backup link, when primary link is failed.

Availability and reliability of this hybrid communication link is better than for separate links. Important part of FSO/RF is switch or router, which makes switching between FSO and RF link. Generally speaking, the availability of the link is given in percent. That means how many percent of the year, link is available. Distance of link influences on overall availability and it is a limitation factor too. For enterprise applications, link availability requirements are greater than 99 %. If FSO systems are used in telecommunications applications, they will meet with higher availability requirements. Carrier class availability is generally considered to be 99,999 % (i.e. five nines) [3]. Thus it is important to back up FSO link with RF link, because in time when factors which have negative effect on FSO signal cause fades of FSO link, switch or router will switch to RF link and communication continues. Availability of this hybrid system is getting close to required five nines.

II. Characteristic of Hybrid FSO/RF Link

The FSO communication link we can define as telecommunication technology which use infrared optical beam to transmit information between two static points. It is a broadband communication technology which need direct visibility between transmitter and receiver. This technology uses optical modulated pulses for fibreless optical data transfer [4]. In free space optics light pulses are transmitted via atmosphere. Due to atmosphere conditions FSO technology has a number of drawbacks. Whenever the medium´s condition is unstable and unpredictable, it becomes difficult to manage the FSO transmissions. The main causes of disruptions are fog, absorption, scattering and scintillation.

Secondary negative influences on FSO link are caused by physical obstructions, buildings sway, rain and snow.

In RF communication link, data is transmitted through the air too, but by digital radio signals at a given frequency, typically in the 9 kHz to 300 GHz range [4]. In combination with FSO systems, RF systems use 60 GHz antennas. Beam width for antenna´s diameter 30,48 cm is 4,7°. This 60 GHz antenna allows using high speed bit rate and connection has high resistance to interference, high security and multiple use of frequency [5]. Factors which have negative effects on FSO are negligible for RF systems. This is a main reason, why are RF links used as backup links.

For FSO systems, fog and with fog join visibility is critical factor for availability and reliability of them. On the other hands, RF links show almost negligible fog attenuation, while they usually suffer from other precipitation types like rain or wet snow. Combining these two technologies into FSO/RF hybrid network may increase overall availability, guaranteeing quality of service and broadband connectivity regardless of atmospheric conditions [6]. The RF link should be available whenever the FSO link is not i.e. it should not be influenced by fog or other weather effects reducing visibility.

III. Visibility and Fog Sensor

As was mentioned above, visibility has major influence on overall availability of hybrid FSO/RF link. Thus it is important to know value of visibility in place where the hybrid link will operate. At the Technical University in Košice we have a device for measuring density of fog, relative humidity and temperature. It is called Fog sensor which was developed like low cost device for experimental purposes (Fig 1.).

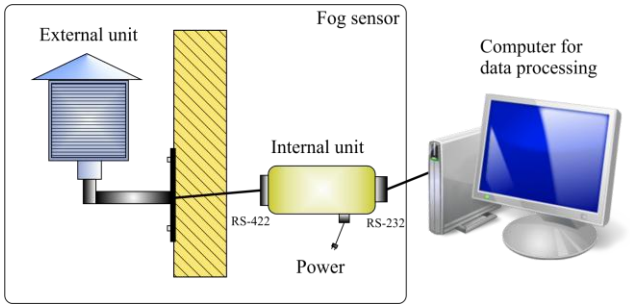


Fig. 1. Hardware scheme of Fog sensor.

These three parameters are important for subsequent steady and statistical evaluation of the quality of FSO/RF communication environment [7].

Fog sensor consists of two main parts, external and internal unit which are connected via 8 pins cable with RS-422 line. External unit does not need external power and it performs the measurement. Internal unit is connected to the computer via serial cable RS-232.

Obtained measured data is processed in PC in simulation program Matlab that allows us to obtain another parameter which describes fog. This parameter is marked like LWC and it means Liquid Water Content of the atmosphere. It describes fog as mass of water drops in the volume units. It is measured in [g/m^3]. From LWC we can calculate a visibility V [km]:

$$V = d \cdot (LWC)^{-0,65} . \quad (1)$$

Parameter d is without unit and takes on specific values for different fog conditions as is shown in TABLE I [8].

TABLE I
TYPICAL VALUE OF PARAMETER "*d*" FOR DIFFERENT TYPES OF FOG

Type of fog	Parameter "*d*"
Dense haze	0,013
Continental fog (dry and cold)	0,034
Maritime fog (wet and warm)	0,060
Dense haze and selective fog	0,017
Stable and evolving fog	0,024
Advection fog	0,02381

In our case, we have mentioned stable and evolving fog. It frequently occurs in central Europe. So visibility is given by following equation [8]:

$$V = 0,024 \cdot (LWC)^{-0,65} . \quad (2)$$

IV. AVAILABILITY OF HYBRID LINK

Availability of FSO link strongly depends on fog and its influence on communication link. In this case fog is represented by visibility. For FSO communication link, there are two possibilities how to obtain information about overall availability. First, we can capture information about fades and then percentual evaluate data during whole year. Another way how to determine availability of FSO link is capture information about visibility in place, where the link will operate. For second way, the fog sensor is a good device for measuring visibility. In this paper, two FSO systems will be used for calculating availability in campus of Technical University in Košice (TUKE). First FSO system is Lightpointe, Flightstrata 155E, second system is FSONA, Sonabeam 155 E. Edge values of visibility for 1 km FSO link is calculated by software package FSO SystSim, which was created at TUKE. This software calculates availability for specific systems by mathematic models and energy balance of FSO connections. After entering inputs parameters of individuals systems, edge values of visibility was calculated. Edge visibility for 1 km long FSO link is 515m for Lighpointe 155E and 620m for Sonabeam 155 E. From this distances is known, that Sonabeam will have worst availability than Lightpointe.

On the other hand, the main drawback of RF link is caused by rain. If we want to determine availability of RF link for 1 km, we need to know how often per year are rainy days. Measuring about fallen precipitations is performed worldwide. With these results of measurement, models of world rainy zones have been proposed (Fig.2). International Telecommunication Union has classified these rainy zones in different regions around the world. From Fig. 2 we can see that Slovakia belongs to rainy zone letter H.

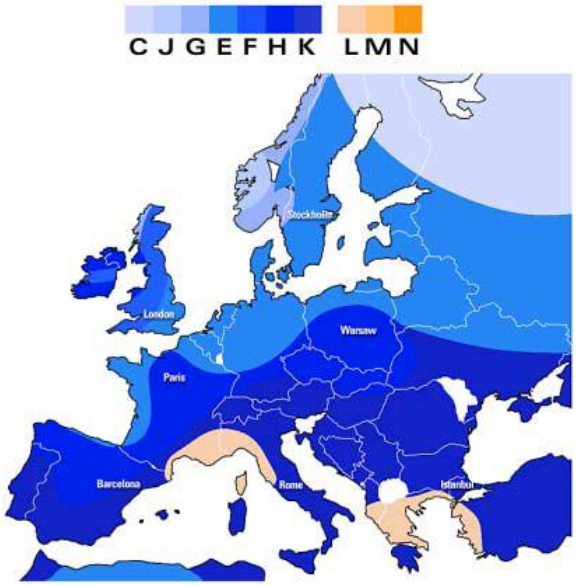


Fig. 2. Rainy zones in Europe.

For 60 GHz RF link 33 cm antenna is used from HXI Company. Availability of 60 GHz link for 1 km is 99,9 % [9]. In these days communication is reliably when its availability is 99,999 %. It means, only 318 seconds per year can be hybrid link down.

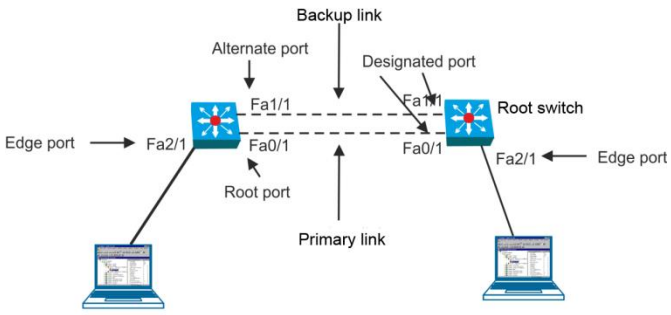


Fig. 3. Hybrid FSO/RF link.

When we join these two communications systems together to hybrid FSO/RF link it is possible to obtain five 9´s

availability (Fig.3). These two links are switching by CISCO switch with RSTP protocol. It is important to connect primary link to lower port numbers on both switches. Time required to switch between links is 6 seconds.

The sample calculation of availability of hybrid link is shown below for year 2010:

$$D_{FSO} = \left(1 - \frac{time\,of\ \ fades}{total\,time}\right) \cdot 100\,\%, \tag{3}$$

$$D_{FSO} = \left(1 - \frac{304}{2 \cdot 8760}\right) \cdot 100 = 98.2648\,\%, \tag{4}$$

where D_{RF} is availability of FSO Flightstrata 155E. Total time is given like two multiply hours per year due to half an hour measuring interval [5].

$$D_{RF} = \left(1 - \frac{time\,of\ \ fades}{total\,time}\right) \cdot 100\%, \tag{5}$$

$$D_{RF} = \left(1 - \frac{304{,}08466}{8760}\right) \cdot 100 = 96.5287\,\%, \tag{6}$$

where D_{RF} is availability of RF link. Total time is given as hours per year. *Time of fades* is given by [5]:

$$time\,of\ \ fades = NS \cdot TS + TF_{FSO}, \tag{7}$$

where *NS* is number of switching RF link, *TS* is total time, TF_{FSO} is time of fades of FSO.

$$time\,of\ \ fades = 51 \cdot 0.00166 + 304\ . \tag{8}$$

Other calculated values for FSO systems and for RF system are entered in the next two tables:

TABLE II
AVAILABILITY OF FSO LINKS

FSO system	Year	Availability
Flightstrata155E	2007	98,1621
Sonabeam 155-E		97,1095
Flightstrata155E	2008	98,3904
Sonabeam 155-E		98,1963
Flightstrata155E	2009	98,5308
Sonabeam 155-E		98,3288
Flightstrata155E	2010	98,2648
Sonabeam 155-E		97,9019
Flightstrata155E	2011	98,2368
Sonabeam 155-E		98,1287

Overall availability of hybrid FSO/RF link using FSO Flightstrata 155E is:

$$D_{FSO/RF} = \left[1 - \left(1 - D_{FSO}\right) \cdot \left(1 - D_{RF}\right)\right] \cdot 100 = \\ = 99{,}99832\,\%. \tag{9}$$

TABLE III
NUMBERS OF SWITCHING RF LINK

Year	Flightstrata 155E	Sonabeam 155-E
2007	48 times	62 times
2008	44 times	46 times
2009	48 times	49 times
2010	51 times	58 times
2011	49 times	53 times

Overall availability of hybrid FSO/RF link using FSO Sonabeam 155-E is 99,99809 %.

V. CONCLUSION

Hybrid FSO/RF link uses both links strength to overcome their weaknesses. By combining these two systems will the hybrid FSO/RF link be able to achieve higher availability of transmission lines than an independent FSO or RF line. Because of the RF hybrid FSO/RF can reduce the impact of heavy fog and FSO can reduce the impact of the rainfall. Dense fog is the most important adverse factor of FSO links and rain is the most important adverse factor of RF connections. Advantages of using FSO link in case of lost connection may not be lost due to using a 60 GHz RF system as backup link which has similar properties than FSO link. Average availability of hybrid FSO/RF system in Košice is 99, 998205 %.

ACKNOWLEDGMENT

We support research activities in Slovakia / Project is co-financed from EU funds. This paper was developed within the Project "Centrum excelentnosti integrovaného výskumu a využitia progresívnych materiálov a technológií v oblasti automobilovej elektroniky", ITMS 26220120055.

REFERENCES

[1] Z. Kolka, O. Wilfert, V. Biolkova "Reliability of Digital FSO links in Europe," Int. J. Electronics, Communications, and Computer Engineering, vol. 1, no. 4, pp. 236–239, 2007.
[2] M. Reymann, J Piasecki, F. Hosein and col. Meteorological Techniques, July 1998.
[3] I. I. Kim, E. Korevaar, "Availability of free space optics (FSO) and hybrid FSO/RF systems," Proc. of SPIE, vol. 4530, pp. 84-95, 2001.
[4] M. Ahmad, A. Fuqaha, O. Awwad, B. Khan, "Synergies of Radio Frequency and Free Space Optics Communication: New Hybrid Solutions for Next Generation Wireless Mesh Networks", International Journal of Computer Networks (IJCN), Vol. 4, 2012.
[5] R. Luna, D. K. Borah, H. Tapse, „Behavior of hybrid optical/RF channels...", Proc. IEEE GLOBECOM, Honolulu, HA, November-December 2009.
[6] D. M. Forin, G. Incerti, G. M. Tosi Beleffi, B. Geiger, E. Leitgeb, F. Nadeem, „Free Space Optical Technologies", Trends in Telecommunications Technologies, March 2010.
[7] Ľ. Ovseník, J. Turán, P. Mišenčík, J. Bitó, L. Csurgai-Horváth, „Fog density measuring system", Acta Electrotechnica et Informatica, 2012, 1-5.
[8] S. Sheikh Muhammad, M. Saleem Awan, A. Rehman. „PDF Estimation and Liquid Water Content Based Attenuation Modeling for Fog in Terrestrial FSO Links". Radioengineering, Vol. 19, No. 2, June 2010.
[9] Characteristics of Precipitation for Propagation Modelling, Recommendation ITU-R PN.837-1.

Video content protection using digital watermarking based on DCT - SVD transformation

[1]*Martin BRODA (1st year),* [2]*Jozef LIPTÁK (2st year)*
Supervisor: [3]*Dušan LEVICKÝ*

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

[1]martin.broda@tuke.sk, [2]jozef.liptak@tuke.sk, [3]dusan.levicky@tuke.sk

***Abstract*— This paper is focused on protection of video data in real time against creating illegal copies. On this purpose can be used digital watermarks. The main idea is providing security of copyright. Main point of this article is creating proposal of method for digital watermarking in video. The specified method is realized in transformed domain and it is based on combination of Discrete Cosine Transformation and Singular Value Decomposition.**

***Keywords*—digital watermarking, video, copyright, transformed area**

I. INTRODUCTION

Nowadays multimedia can be shared over Internet very easily, what may have to consequence propagation illegal copies of multimedia.

Embedding invisible digital watermark into video is one facility, how multimedia can be protected against illegal copying. The objective of this article is design of method for digital watermarking in video in real time. Basic scheme is based on application of Discrete Cosine Transform (DCT) and Singular Value Decomposition (SVD). This method should provide good ratio between perceptual transparency and robustness against different types of attacks.

II. DIGITAL WATERMARK

A. Theory of watermarks

Digital watermarking is one from the most effective method, which is used as a solution to copyright problem. Digital watermark is represented as a code that uniquely identify author and is embedded into digital data. Watermarks can be represented in form of symbols, images, logos, short forms of speech or binary information. [1]

Watermarks can be extracted from data in law action, whereby we can proof the origin of this data. It is realized so, that extracted watermark is compared with the watermark, which was embedded into original data.

B. Basic requests on watermarks

Watermarks should satisfy these six requests :

- **Nonperceptibility** – is a very important parameter for digital watermarks. Watermark is considered as nonperceptibility, when a human with his senses can´t recognize difference between original and watermarked data.
- **Robustness** – expresses difficulty of watermark removal from data without knowledge of the method that was used for embedding watermark. This characteristic can be also realized than robustness on changes and modifications original data (e.g. compression, filtering, cropping and etc.)
- **Security** – is based on characteristic, that potential attacker don´t know algorithm for embedding watermark, which was used. Security can be increased using a one or few cryptographic keys, which we realize encryption of the embedded watermark.
- **Undetectability** – defines equality between original and watermarked data in term of source information.
- **Capacity –** specifies size of watermark, which can be embedded into original data.
- **Complexity –** explains effort that we need to invest on detection information about watermark.

The most important requests are nonperceptibility, robustness and capacity [2]. Nowadays, we don't have method for digital watermarking that verifies all three requests. Consequently we need to search a compromise between requests.

C. Methods for embedding digital watermarks

Actual method for digital watermarking can be divided into three categories according to area for embedding watermark [3].

Spatial domain – This method realizes to modify image or audio data straightly embedding watermark in spatial (image) domain. The best-known method in this domain is LSB method.

Transform (frequency) domain – is realized to modify transform coefficient of original data by transform coefficient of watermark [3]. The most used transformations are Discrete Cosine Transform DCT or Discrete Wavelet Transform DWT. After embedding watermark we recover watermarked data using inverse transformation.

Parametric domain – Methods in this domain are realized

to change parameters of original signal. Firstly, we must convert data into parametric area. For example, this category contains fractal coding images.

III. DIGITAL WATERMARKING IN VIDEO

Video is very popular media in nowadays. Video data are often mixed with audio in order to create multimedia data.

In history, there was used analog form of video, but now we often have video from camera in defined compression standard. Digital video is also converted from analog video and is represented by a sequence of frames. There is used characteristic (deficiency) of human vision, that continuous video get to playing a lot of frames after themselves. Every frame is composed from pixels. The total number is given as a multiply number of pixels in height and width of the image. Every from this pixels is described level of luminance or color. Digital video provide good characteristics in its next processing, for example digital watermarking.

Methods for digital watermarking in video can be divided into three categories [4]:

- *algorithm based on information from watermarking static image,*
- *algorithm based on principle time domain of video,*
- *algorithm based on compressing video.*

In first category, video is divided into sequence of frames and watermark is embedded into every of them. This method is characterized to longer time for embedding watermark, because video consists from several thousand frames. This algorithm does not exploit characteristics and parameters of video, so that there can be used all known methods for watermarking static images.

Algorithms in second category uses it, that video is dynamic medium and its content and parameters are dependent in time. Methods based on time dimension should have higher capacity for embedding watermark. Watermarking algorithms are divided by time dimension into three categories [4]:

- video as 1D signal,
- video as time signal,
- video as 3D signal.

Algorithms based on video compression use information about compression standard in video. Watermark is embedded directly into compression stream in order to modify this stream. Algorithms in this category do not remove all redundant data and then this data are used for embedding watermark. This method contains three subcategories [4]:

- *methods, which influence coefficients in transform domain indirectly,*
- *methods, which modify information about motion,*
- *methods, which influence directly code words VLC.*

IV. REALIZATION SCHEME FOR DIGITAL WATERMARKING IN VIDEO

This scheme for watermarking is realized using combination DCT and SVD. This transformation was selected following knowledge and experiences with their application in other methods for watermarking, where this transformation achieved very good characteristic and robustness against different types of attacks.

A. Discrete cosine transformation

Watermarking methods often contain DCT (1), because this transformation is used in compression image (JPEG) or video (MPEG). DCT coefficients for block *NxN* are calculated by formula [5]:

$$F(i,j)=\frac{2}{N}C(i)C(j)\sum_{i=0}^{N}\sum_{j=0}^{N}f(x,y)\cos(\frac{(2x+1)i\pi}{2N})\cos(\frac{(2y+1)j\pi}{2N}), \quad (1)$$

Where *f(x,y)* are particular image pixels, *F(i,j)* are transform coefficients. *C(i), C(j)* are constant and they are defined by formula:

$$C(x)=\begin{cases}\frac{1}{\sqrt{2}} & if \quad x=0\\ 1 & in \quad other \quad cases\end{cases} \quad (2)$$

Embedding watermark into data is realized to modify transform coefficient of data with transform coefficients of watermark. Watermarked data are acquired application inverse DCT transform on modify transform coefficient of data.

Major part of image information is contained in low-frequency coefficients. Coefficient, which contains the most important information about image is called DC (direct current) coefficient. This coefficient is situated in upper left in block of transform coefficients. Other coefficients with higher frequency are called AC (alternate current).

B. Singular Value Decomposition

SVD is new method and it is used in mathematic operations with matrixes or for searching in a large amount of data. SVD (3) defines that matrix with real numbers can be divided on conjunction three matrix with same sizes than original matrix. Decomposition original matrix accelerates calculation, because nowadays we work with large size matrixes [6].

When A is square matrix, then singular value decomposition may be described by formula [6]:

$$A=USV^{T}, \quad (3)$$

A is original matrix, *S* is a diagonal matrix of singular values, *U* is orthogonal matrix with left singular vector and *V* is orthogonal matrix with right singular vector. Diagonal matrix *S* with singular values has a form as follows:

$$S=\begin{pmatrix}s_1 & 0 & 0 & 0\\ 0 & s_2 & 0 & 0\\ 0 & 0 & \dots & 0\\ 0 & 0 & 0 & s_n\end{pmatrix}, \quad (4)$$

where elements *s* on diagonal are singular values, which satisfy condition:

$$s_1 \geq s_2 \geq \dots \geq s_n, \quad (5)$$

C. Scheme for embedding watermark using DCT-SVD

Scheme for embedding watermark is illustrated on Fig. 1 and contains operations as loading video (divided into frames), convert RGB color model to YC_bC_r, realization of DCT and SVD, and finally embedding watermark into finite number frames of video.

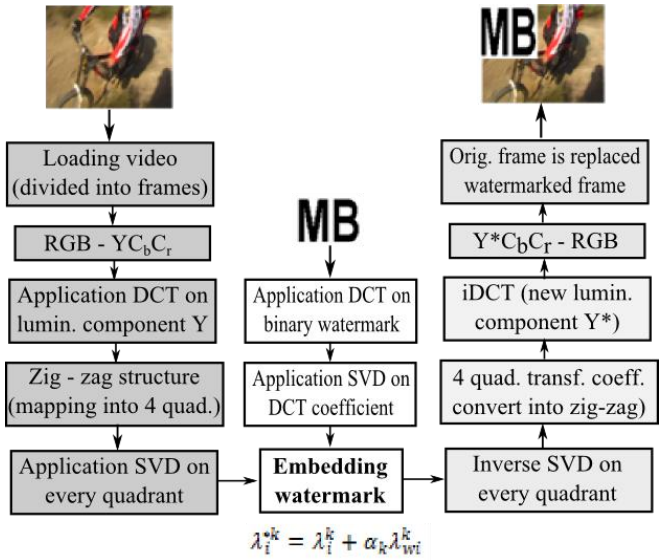


Fig. 1 Block diagram for embedding of watermark

In first step, we load video and it is divided into frames. Watermark is only embedded into finite number of frames (watermark was embedded into three frames for testing). Next, frame is converted from RGB to YC_bC_r model. In designed method watermark is embedded into luminance component, whereby is secured robustness against attacks, for example against compression.

Next operation is application DCT on luminance component Y of frame. Zig-zag structure of DCT coefficients is converted into 4 quadrants, than it is illustrated on Fig. 2.

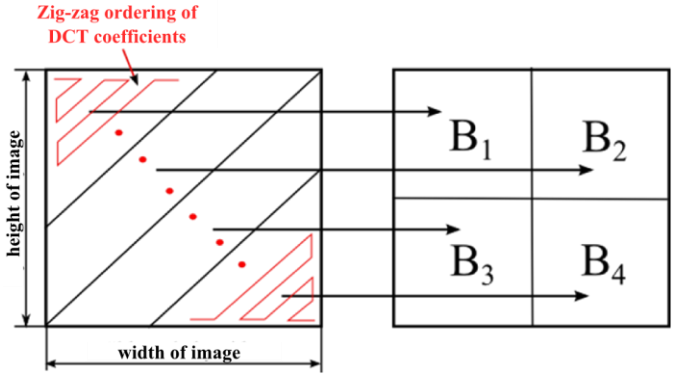


Fig. 2 Mapping DCT coefficients into four quadrants

In next step on every quadrant is used SVD transform in order to acquiring *U*, *V* and *S* matrixes. Watermark is represented by binary images (for example image with letters MB). DCT and SVD transform is also realized on this binary watermark, whereby there were gained matrixes U_W, V_W and S_W for watermark. Watermark is embedded as modification singular values of image using singular values of watermark (6). This operation can be defined by formula:

$$\lambda_i^{*k} = \lambda_i^k + \alpha_k \lambda_{wi}^k, \qquad i=1, 2, \ldots, n \tag{6}$$

Where λ_i^{*k} are modified singular values from matrix *S** in quadrant k, λ_i^k are singular values (matrix *S*) of original frame in the quadrant *k*, λ_{wi}^k represents singular values of watermark (matrix S_W) and α_k are scaling factors for every quadrant *k*. Scaling factor influences to strength of modification singular values. Better robustness of watermarking is caused by higher value of scaling factor, but on the other side it can cause distortion of frame.

Next, there are inverse transformations or operations that are illustrated in Fig. 1. Result of this all operations is a watermarked frame.

D. Scheme for extracting watermark using DCT-SVD

This part is based on extracting watermarks from frames in video. Process of extraction can be described by this formula:

$$\lambda_{wextrakcia i}^k = (\lambda_i^{*k} - \lambda_i^k) / \alpha_k \tag{7}$$

Where $\lambda_{wextrakcia i}^k$ are singular values of extracted watermark in every quarter. This value is acquired difference between singular values of watermarked image and singular values of original image. This difference is divided by scaling factor with the same size as in process embedding watermark.

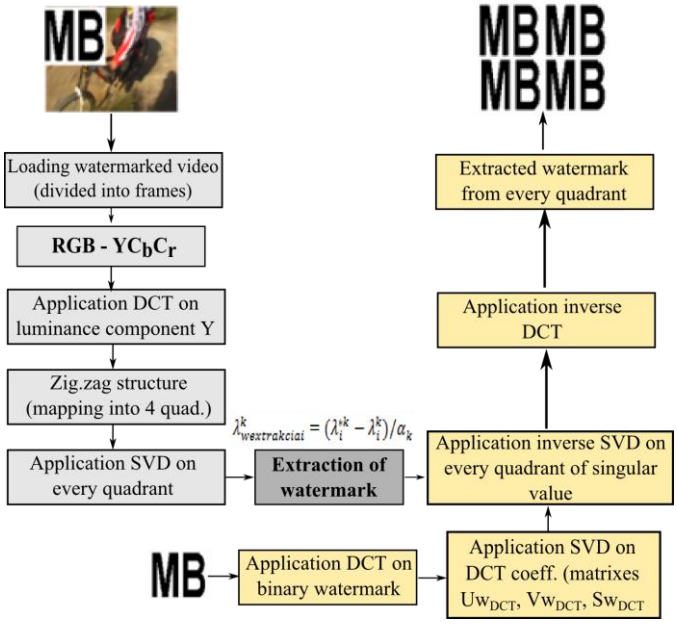


Fig. 3 Block diagram for extracting of watermark

Fig. 3 shows blocking scheme for extracting watermarks. We get four watermarks, because we realized extraction from every quadrant. For extracting of watermark is necessary to have original and watermarked image.

V. TESTING

Videos for testing were divided into three categories in term of their dynamics. A dynamic video is characterized by fast motion of objects, so that adjacent frames are more different. On other side, there are videos with smaller dynamic, where adjacent frames of video are similar. Resolution of these videos is in QCIF format (352x288 pixels) and watermark is binary image with size 176x144 pixels. Size of watermark is defined by watermarking method, because watermark is embedded into every quarter of frame.

Influence of embedding watermark on video quality is measured using objective parameters *MSE* (Mean Square Error) and *PSNR* (Peak Signal Noise to Ratio). Calculated values of these parameters for different size of scaling factor are illustrated in Tab. 1. Growing value of scaling factor (α_1 is

scaling factor for first quadrant of image and $\alpha_{2\text{-}4}$ are scaling factor for 2-4 quadrant of image) causes higher degradation quality of picture. Picture impairment results in decrease value of *PSNR* (dB) and on the other side it causes increase of *MSE* value.

Tab. 1 Comparison MSE and PSNR for different types of videos and for the different scaling factors

Scaling factor		Type of video		
		dynamic	medium dynamic	little dynamic
α_1=2, $\alpha_{2\text{-}4}$=0.8	PSNR [dB]	52,227	51,734	52,430
	MSE	0,389	0,436	0,374
α_1=5, $\alpha_{2\text{-}4}$=1.5	PSNR [dB]	47,212	46,528	47,414
	MSE	1,236	1,447	1,180
α_1=10, $\alpha_{2\text{-}4}$=3	PSNR [dB]	41,662	40,880	41,777
	MSE	4,431	5,310	4,320
α_1=20, $\alpha_{2\text{-}4}$=4	PSNR [dB]	37,842	37,023	37,970
	MSE	10,688	12,905	10,377

Robustness of watermarking method was tested against intended and unintended type of attacks. Unintended attack was realized using lossy compression standard VP8 in video for different scaling factor and different types of videos (Tab. 2). The quality of extracted watermark is defined using bit match *BM* (%) between original and extracted watermark.

Tab. 2 The quality of the extracted watermark for VP8 compression

Type of video	Number of quarter (extracting watermark)	Scaling factor		
		α_1=2, $\alpha_{2\text{-}4}$=0.8	α_1=5, $\alpha_{2\text{-}4}$=1.5	α_1=10, $\alpha_{2\text{-}4}$=3
		BM [%]	BM [%]	BM [%]
dynamic	1.	93,79	99,87	100
	2.	85,16	100	99,92
	3.	90,59	93,65	96,65
	4.	75,87	81,38	85,36
medium dynamic	1.	99,86	100	100
	2.	99,98	100	100
	3.	99,9	99,95	99,87
	4.	90,12	96,74	98,84
little dynamic	1.	99,99	100	100
	2.	100	100	100
	3.	99,71	100	100
	4.	95,65	99,17	99,45

Next, we tested influence to one of intended attacks, specifically attacks realized by frame averaging in video. This attacks were realized on watermarked videos with values of scaling factor α_1=5, $\alpha_{2\text{-}4}$=1.5. We calculate average from 2, 3, 5 and 10 frames from the same scene of video. Increase number of frame for averaging can cause higher level of picture degradation, especially for dynamic video, because adjacent frames are very different. Calculated values for bit match are in Tab. 3.

Tab. 3 Testing the quality of the extracted watermark for attack based on frame averaging

Type of video	Number of quarter (extracting watermark)	Number of frame averaging			
		2 frames	3 frames	5 frames	10 frames
		BM [%]	BM [%]	BM [%]	BM [%]
dynamic	1.	89,03	87,87	86,08	81,43
	2.	88,45	86,41	84,64	83,5
	3.	90,59	87,71	83,94	79,66
	4.	97,64	95,38	90,59	83,21
medium dynamic	1.	87,2	82,73	80,17	79,7
	2.	96,3	94,26	92,91	89,43
	3.	99,08	97,71	95,27	90,76
	4.	99,73	98,66	95,27	86,27
little dynamic	1.	86,25	86,01	85,68	85,48
	2.	83,9	82,88	82,07	81,57
	3.	85,9	83,83	82,05	80,31
	4.	94,12	92,05	89,58	85,77

VI. Conclusion

This paper provides basic information about methods for digital watermarking in video and specifies watermarking method based on combination DCT-SVD transformations. Experimental results show, that this watermarking method is very robust against video compression, for example VP8 codec. Bit matches for unintended attack based on VP8 compression were from 99 to 100 percent at least in one from 4 extracted watermarks.

Robustness against intended attack based on frame averaging in video was also tested. Bit matches for watermark achieved values over 80 percent also for averaging ten video frames.

Disadvantages this method are higher computing power and higher time consumption in the case, that we need to watermark every frame in the video.

Acknowledgment

The work presented in this paper was supported by Ministry of Education of Slovak Republic VEGA Grant No. 1/0386/12.

References

[1] D. Levický, *Multimediálne telekomunikácie.* Košice: Elfa s.r.o., 2002. 240 s. ISBN 80-89066-58-5.

[2] M. Wu – B. Liu, *Multimedia data hiding*, Springer-verlag New York, Inc. 2003.

[3] D. Levický – R. Hovančák – Z. Klenovičová, *Technika digitálnych vodoznakov. Princípy, systémy a použitie.* In Slaboproudý obzor, Roč. 59, Číslo 1, pp. 1-5, 2002.

[4] G. Doerr, *Security Issue and Collusion Attacks in Video Watermarking*. PhD Thesis, Universite De Nice Sophia – Antipolis, 2005.

[5] M. Arnold – S. Schmucker – D. Wolthusen, *Techniques and Applications of Digital Watermarking and Content Protection.* Norwood: Artech House, inc., 2003. 274 s. ISBN 1-58053-111-3.

[6] L. Rajab – R. Al-Khatib – A. Al-Haj, *Video watermarking Alghoritmus Using the SVD Transform*, European Journal of Scientific Research Vol.30 No.3, pp.389-401 2009.

Virtual differential protection

[1]*Marian HALAJ (5st year)*, [2]*Peter HERETIK (2st year)*
Supervisor: [3]*Ladislav VARGA*, [4]*Justin MURIN*

[1,3]Dept. of Electrical Power Engineering, FEI TU of Košice, Slovak Republic
[2,4]Institute of Power and Applied Electrical Engineering, FEI STU of Bratislava, Slovak Republic

[1]marian.halaj@centrum.sk, [2]peterheretik@gmail.com, [3]ladislav.varga@tuke.sk, [4]justin.murin@stuba.sk

***Abstract*—Differential protection is the fastest, but also most difficult and expensive way to protect the busbar against the short circuit. Published paper shows innovative ways to design new relays.**

Keywords*—Differential protection, IEC *61850

I. INTRODUCTION

Faults on busbars are considered as one of the most critical defects on the power distribution systems. Very high short circuit currents sometimes of hundreds of kilo-amps may cause very expensive damages on the equipment, sometimes even the risk for personnel. Such faults must be cleared within few milliseconds.

In general we recognize three concepts of busbar protection:

- Differential protection (87B)
- Overcurrent protection (51)
- Directional overcurrent protection (67)

Main task of each busbar protection is to make sure the proper operation of the power system protection is to identify the zone where the fault occurred and keep the selectivity of tripping. This condition is important from the operation point of view, because in same cases the lost of service may cause much higher losses than the fault itself. Only if the breaker of the faulty section is not able to isolate this section, than the upstream breaker will trip.

II. BUSBAR DIFFERENTIAL PROTECTION

Differential protection is the fastest type of power system protection, therefore is mainly considered as prime protection of important busbars. Relay operates without time delay within shortest available time, usually 40-50ms.

Busbar differential protection basically operates on the differential current principle. The relay measures balance between the current flowing into the busbar normally from incoming feeder(s) and summary of currents flowing to the outgoing feeders. If the zero balance is not kept, the fault has occurred on busbar. This is the very simple explanation that doesn`t consider the influence of coupling breakers. Bus coupler may either supply busbar the current or behave as a consumer depending on bus configuration. Another aspect is that the current are compared separately for each phase. Also the construction of this system is very complicated and sensitive. Special attention shall be taken to the electro-magnetic compatibility issues (EMC) like pulses in the network. Proper settings of the relay shall block the operation of the relay under the following conditions:

- Faults occurred out of protected zone
- Non symmetrical load on the phases
- Harmonics in the network
- Second harmonic during switching of transformers
- Transient events

To ensure the proper operation of the relay under above conditions the proper network studies and safe margins inside relay shall be performed.

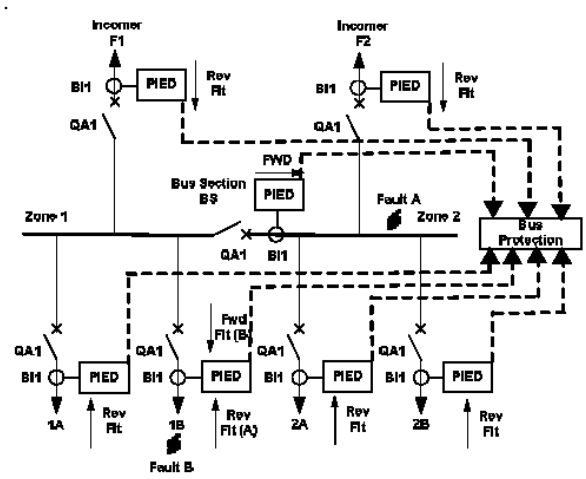


FIGURE I
DIFFERENTIAL PROTECTION PHILOSOPHY

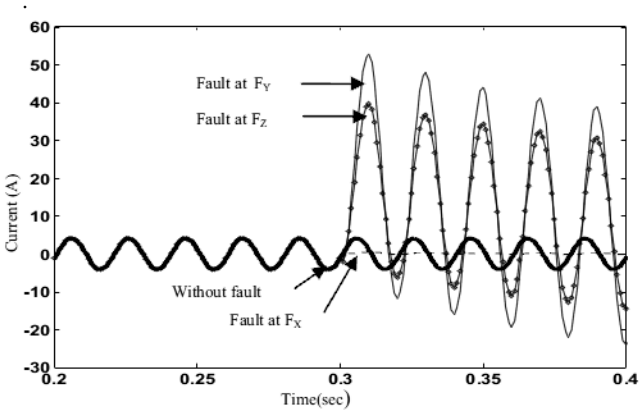


FIGURE II
SHORT CIRCUIT CURRENT ON BUSBAR

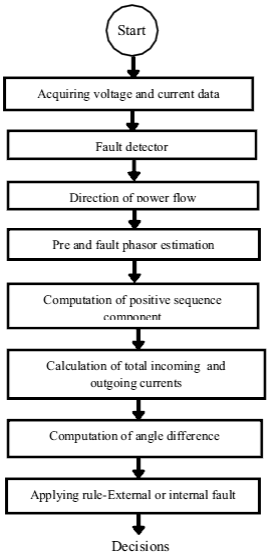


FIGURE III
OPERATION ALGORITM OF DIFFERENTIAL PROTECTION

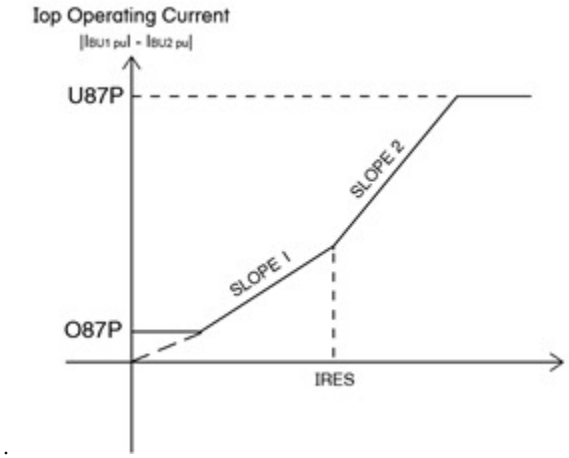


FIGURE IV
TRIPPING CHARACTERISTIC OF DIFFERENTIAL RELAY

Typical connection of the busbar differential protection is shown on (Figure I). Merging units measure the current in each feeder and bus coupler. Current is measured by current transformers connected to the analogue inputs of merging unit (Figure V). Current is sampled inside merging unit with the frequency of 80 samples / cycle and consequently transferred to the relay performing differential protection. Another way is the hard wired connection of all current transformers to the binary input cards of protective relay. Differential protective relay is microprocessor based relay (IED) with the configured logic (Figure III). Configuration file imports all required analogue and binary inputs from MUs. The most important inputs are current in each feeder, breaker position, watch dogs and trip circuit supervision relays. MU transforms the current information into the voltage one and makes the sampling procedure. Configuration file shall enable several stages of tripping characteristic as well as compensation of inputs if different types of current transformers are used. In case of fault on busbar the relay sends opening order to all circuit breaker via hard-wired connection according to tripping characteristic (Figure IV).

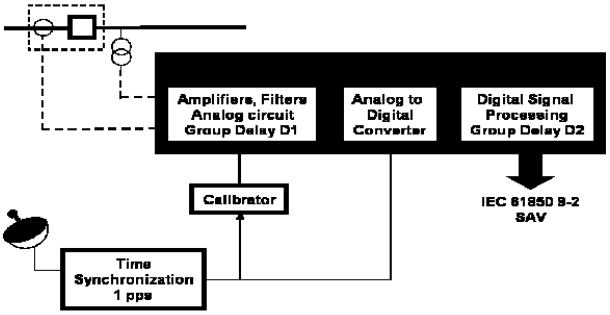


FIGURE V
MERGING UNIT (MU)

III. IEC 61850 PROTOCOL

IEC 61850 was created to be an internationally standardized method of communication and integration and to support systems built from multivendor intelligent electronic devices (IEDs) that are networked to perform protection, monitoring, automation, metering, and control [1]. This protocol enables wild range of information exchange between different IEDs, like measurements, status indication, interlocking, closing and opening orders, etc.

From the functional point of view the IEC 618500 operates on three different levels: process, bay and station level [4].

Process bus facilitates the time critical communication between protection and control IED to the process (the primary equipment in the substation), such as sampled values, binary status signals or binary control signals. This layer of the substation is related to the gathering information, such as voltage, current, status indication, etc. Data transfer within this layer is available in two ways. Unidirectional multi-drop point-to-point fixed link carry a fixed dataset of the length of 16 bites. This is appropriate for transfer of sampled values. Sampled measured values (SMV) data transfer is able to transfer the values of different sizes as well as process that together.

Station bus facilitates communication between station level and bay level. Station bus provides primary communication between different logical notes. This communication includes also protection functions, control and supervision. Communication is available on connection oriented bases or GOOSE.

A. Generic Object Oriented Substation Event (GOOSE)

GOOSE communication is one of the main features of IEC 61850 [5]. It enables data exchange on station bus based on the publisher-subscriber concept. Comparing to another station bus protocol connection oriented base the GOOSE message may be received by several receivers. This feature was already available in older communication protocols using client-server approach as well, but in that case the communication worked on all 7 layers that is significantly time consuming. GOOSE message principle doesn't confirm the delivery receipt, but repeats the message several times instead. This feature provides very fast communication. Messages can be categorized according to importance by means of SCL. Most critical messages, like tripping or blocking signals may be received within 3ms that is in certain cases faster than hard-wired connection. Fast acting functionality of these messages enables us to model the clearing of most critical faults on electrical power

distribution, like short circuits on busbar.

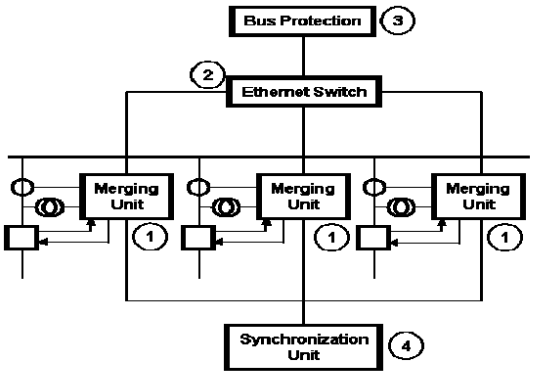


FIGURE VI
PROCESS BUS COMMUNICATION

IV. MODEL OF VIRTUAL PROTECTION

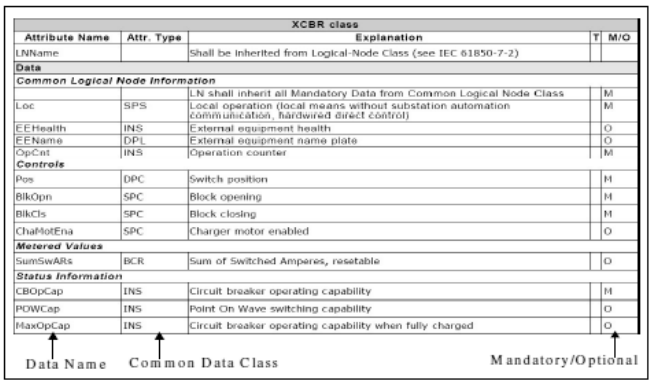
Process stage is similar to the one that makes conventional busbar differential protection. Current information is taken from current transformers of class 5P or class X that are properly sized. Calculation shall be performed for CT sizing to prevent of incorrect current measuring. In addition to saturation by DC component the calculation shall consider the impact of connected loads, length and cross section of connecting wires. Status information like circuit breaker position, auxiliary supply availability or trip circuit supervision are usually connected tu the MU's binary input module. Those modules are voltage free relay contacts normall energized by low voltage DC or AC voltage of 110 or 230V. After the sampling of the inputs those are published on the process bus. Merging unit is connected to the system via converter and multimode fibre optic (100Mb or 1Gb) that enables simultaneous transfer of plenty information via one cable.

Main advantage of this IEC 61850 platform based system is the horizontal communication on the publisher / subscribers principle that enables communication of the devices on the same level. This is completely different principle then conventional master-slave protection. Data exchange between different feeders of the same switchgear is available. One of the results of this fact is that there is no need for any separate control unit providing differential protection itself. As the data may be exchanged between feeders of the switchgear the differential protection may work on any of them. The advantage is a backup of the master relay or possible service / maintenance activity on the energized switchgear.

Data exchange between the merging units (IEDs) is performed via programmed GOOSE messages with the guaranteed transmission speed 3-5ms depending on network traffic. Concept of the differential protection may be generally customized. Since the messages are designed as not confirmed in our case, it allows very fast transfer of the sampled data via network. GOOSE message is not only used for the transfer of measured values and status information. One of the main advantages is the sending of circuit breaker opening and closing contacts. This will minimize the number of hard-wired connections in the network. Tripping signal message takes normally 2,5-3ms so it is even faster then hard-wired connection with relays (approx. 10ms). Using the ring configuration of the fibre optic network we can even improve the reliability of the system. IED's are connected in a loop (ring) configuration. Such a configuration allows proper the operation even in case of broken fibre optic cable and the diagnostic may be done easily. This is not possible by using hard wiring connections.

Station configuration language brings new era into the settings of protection and communication. Lot of operation scenarios may be configured depending on actual operation conditions. By means of the language the algoritm of protection and relations between inputs and outputs are set up. For our purposes we define 4 categories of GOOSE messages. Category one defines the most critical data that is current sample and tripping order. Signals are to be received by subscribers within 2,5-3ms. Second level of messages includes interlocking signals, status indications and alarms with the delivery 15-20ms. Third class is configurable starting with minimum delivery time of 40ms for control and indication purposes and the type four is reserved for the signals with low importance. Categorization of the messages improves the reliability of the system. It enables to deliver critical messages on time without unnecessary delay caused by system overload.

Functionality of the protection is described by logical nodes called 'FEEDER' connected to the logical device gateway placed inside one of the switchgear cubicle. This device is a switch connecting the logical nodes on same horizontal level each other and provides the subsystem connection to the station bus. In fact the logical nodes represent the software inside physical devices, merging units called also IEDs. Logical node of outgoing feeder is



XCBR class

Attribute Name	Attr. Type	Explanation	T	M/O
LNName		Shall be inherited from Logical-Node Class (see IEC 61850-7-2)		
Data				
Common Logical Node Information				
		LN shall inherit all Mandatory Data from Common Logical Node Class		M
Loc	SPS	Local operation (local means without substation automation communication, hardwired direct control)		M
EEHealth	INS	External equipment health		O
EEName	DPL	External equipment name plate		O
OpCnt	INS	Operation counter		M
Controls				
Pos	DPC	Switch position		M
BlkOpn	SPC	Block opening		M
BlkCls	SPC	Block closing		M
ChaMotEna	SPC	Charger motor enabled		O
Metered Values				
SumSwARs	BCR	Sum of Switched Amperes, resetable		O
Status Information				
CBOpCap	INS	Circuit breaker operating capability		M
POWCap	INS	Point On Wave switching capability		O
MaxOpCap	INS	Circuit breaker operating capability when fully charged		O

FIGURE VII
ANATOMY OF LOGICAL NODE

structured into several groups: circuit breaker status indication, earthing switch information, phase and line current and voltage, group of signals for alarms, indication, other metering and control (Figure VII).

V. CONCLUSION

Virtual differential protection enables a cheap alternative to the conventional one. Required function may be performed by any of feeder protective relays or by several relays as a back up. Solution is designed mainly for low voltage high current installations, but it may be installed on medium voltage switchgears as well.

Engineers appreciate the minimizing of the wiring connection and related mistakes during design and testing.

Configuration of the system is much easier comparing to

relay hard wiring and older protocols because many configuration or data change functions are already standardized by this protocol. On the other hand structure of the software and communication may be customized according to the special application.

Operation speed of the system may be customized by choosing of GOOSE communication message that are available with or without return receipt. Ring communication loop improves the reliability comparing to star connection of conventional system.

References

[1] R. E. Mackiewitz, "Overview of IEC 61850 and Benefits". 2006. Available on internet: <http://morse.colorado.edu/~tlen5830/ho /Mackiewicz06IEC61850.pdf >

[2] S. B. Vicente and team, "New protection scheme based on IEC 61850".

[3] S. S. Tarlochan, "Numerical busbar protection: benefits of numerical technology and IEC 61850 " in *Relay Protection and Substation Automation of Modern Power Systems*. Cheboksary, 2007, pp. 1–10.

[4] D. Tholomier - H. Grasset – M. Stockton, "Implementation Issues with IEC 61850 Based Substation Automation Systems " in *Fifteenth National Power Systems Conference (NPSC)*. Bombay, 2008, pp. 473–478.

[5] I. Mesmaeker, "How to use IEC 61850 in protection and automation". Available on internet: <http://www05.abb.com/global/scot/scot221.nsf /veritydisplay/fed26a71538479c3c12570d50034fbe4/$file/Rapport.pd>

Author's index

13th 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, 448 pages, 2013
Number of CD Proceedings: 150 pieces

Editors: Prof. Ing. Alena Pietriková, CSc.
Ing. Milan Nosáľ
Ing. Ivan Halupka
Ing. Emília Pietriková

ISBN 978-80-553-1422-8