VSB-Technical University of Ostrava
University of Bielsko-Biała

Proceedings of the 11th International Conference

Knowledge in Telecommunication Technologies and Optics

KTTO 2011

June 22-24, 2011
Szczyrk, POLAND

Editors: Miroslav VOZNAK, Jan SKAPA
Igor Piotr KURYTNIK, Bohdan BOROWIK

Knowledge in Telecommunication Technologies and Optics 2011 - KTTO 2011

VSB-Technical University of Ostrava,
17. listopadu 15, 708 33 Ostrava-Poruba, Czech Republic
Department of Telecommunications, Faculty of Electrical Engineering and Computer Science

Authors: collective of authors

First published: Ostrava, 2011, 1st edition
Page count: 242
Publisher: VSB-Technical University of Ostrava
Printed by: VSB-Technical University of Ostrava
Impression: 56 pieces

Not for sale

ISBN 978-80-248-2399-7

Preface

The KTTO 2011 Conference was held in Szczyrk, Poland from June 22nd to 24th, 2011. The authors of about fifty papers, which passed through a blind review process and were accepted, were invited to present them. Two rectors, from University of Bielsko-Biala and Ternopil Ivan Puluj National Technical University, attended the opening ceremony and welcomed more than 60 participants. After their greetings, distinguished persons of Bielsko-Biala town hall and directors of few companies supporting the conference, spoke to the participants with wishes for success of the conference as well. Immediately after the welcome ceremony, invited lectures delivered their speeches in plenary session, next subsequent lectures were divided into eight sessions.

The individual parts of the conference program were moderated by their respective chairman who also managed discussions on presented topics and nominated the best papers in their chair session reports. KTTO 2011 Conference was terminated at University of Bielsko-Biala where the last session took place. The pleasant atmosphere accompanied the conference and all participants contributed to success of the conference by their interesting papers and friendly mood. The social evening, taking place in premises of hotel Orlie Gniazdo, was no less successful.

In conclusion, I can say the main ideas of the conference were fulfilled, this event contributed to networking of researchers and sharing their recent results in field of Telecommunications. I would like to thank all lecturers, authors of papers and their reviewers. I would also like to thank my colleagues from conference committee who actively participated in the preparations and made the success of the conference possible. Last but not least, I would also like to thank our partners and sponsors.

Miroslav VOZNAK
Editor in chief

Table of Contents

Network Technologies and Services

chairmen:

Miroslav VOZNAK and Vladislav SKORPIL

Negative Influences of Branch Lines on xDSL Technologies in the Access Networks

Jan Cuchran, Rastislav Roka
Institute of Telecommunications
Slovak University of Technology
Bratislava, Slovakia
jan.cuchran@stuba.sk, rastislav.roka@stuba.sk

Abstract **- The contribution focuses on a possibility for the multimedia service exploitation in the access networks by means of new xDSL technologies on metallic symmetric lines. In the main part, arrangements for eliminating of negative environmental influences in homogeneous lines are presented.**

Keywords **– xDSL technologies, branch lines, the channel capacity**

I. INTRODUCTION

xDSL ("x" Digital Subscriber Line) [1] – [3] technologies help to solve the problem of insufficient transmission rates in the access network (that provides the connection between subscribers and the local central office) using the most recent approaches and techniques in various areas. The xDSL technologies are technologies of the high-speed signal transport in the environment of symmetric homogeneous metallic lines. They utilize already extant metallic access networks - in the case of the majority of users, there is only one copper pair that provides the limited connection from the residence to the network. They also allow the fast, risk less and efficient exploitation of multimedia services for as many as possible users. For the correct working activity of xDSL equipments on metallic subscriber lines, however, it is also necessary to know limiting conditions of the xDSL technologies utilization for the multimedia service exploitation and to know methods of minimizing or eliminating negative environmental influences. Some methods for modeling of the negative environmental influences can be found in [4] – [8].

II. THE PROCEDURE AT THE REALIZATION OF SUBSCRIBER CONNECTIONS BY MEANS OF XDSL TECHNOLOGIES

On the basis of acquired experiences with the realization of the xDSL connection [9], we recommend to perform the following arrangements before the exploitation of the xDSL technologies into the access networks:

a) To measure the basic cable parameters, i.e. characteristic impedance, DC resistance, propagation loss, signal-noise ratio, near-end crosstalk, and to find out whether the results of these measurements are convenient for international standards, norms and recommendations.
b) To compare the acquired parameters of metallic lines and the recommended parameters introduced by the xDSL modem manufacturers.
c) To check up the whole transmission path, to measure the section parameters and to remove all possible defect occurrences on the transmission path.
d) To minimize the number of branch offices between xDSL modems, the optimal number is zero.
e) In the case that another telephone device on the subscriber side is link up to the line between xDSL modems (Fig. 1), the signal transport in the high-frequency area can be threatened [10].

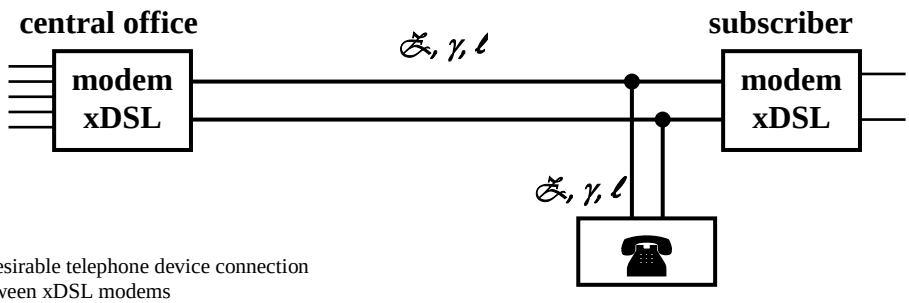


Figure 1. The undesirable telephone device connection between xDSL modems

Warrant: From a viewpoint of the input impedance $\underline{Z}_{input}$, this branch line connection seems to be practically no-load terminated because the branch line length l_1 is around tens of meters (Figure 1), the line attenuation $a = \alpha . l_1 \approx 0$ dB and the terminating impedance $\underline{Z}_2$ closes to the ∞ for frequency bands utilized by xDSL technologies $(\omega >> \omega_r)$ (Fig. 2).

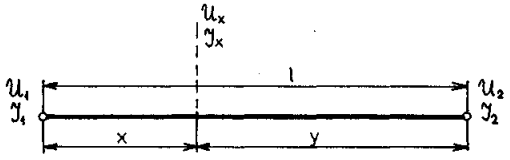


Figure 3. The relationships between *l*, *x* and *y*

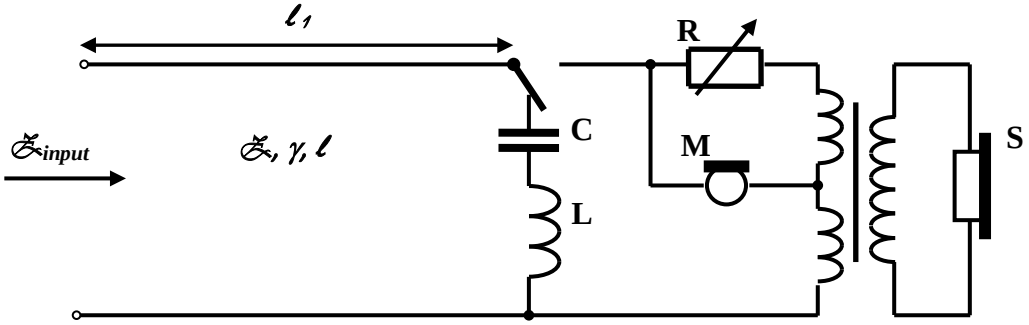


Figure 2. The input impedance of the undesirable telephone device connection

By Fig. 2, $\underline{Z}_2 = j.(\omega.L - 1/(\omega.C))$. For $\omega >> \omega_r$ (in the case of xDSL technologies $f = 6$ MHz, $\omega = 2\pi.6.10^6$), the value of the terminating impedance of the branch line $\underline{Z}_2$ is much larger than the value of the characteristic impedance of the branch line $\underline{Z}$ ($\underline{Z}_2 >> \underline{Z}$). According to [10], $y = l - x$ assigns the arbitrary place from the end of the line (Fig. 3) and for the branch line that behaves as a no-load-circuit line with no losses (Fig. 4) is valid:

$$\underline{U}_y = \underline{U}_2.\cosh\gamma.y + \underline{Z}.\underline{I}_2.\sinh\gamma.y \quad (1)$$

$$\underline{I}_y = \underline{I}_2.\cosh\gamma.y + \frac{\underline{U}_2}{\underline{Z}}.\sinh\gamma.y \quad (2)$$

When $y = l$, then $\underline{U}_y = \underline{U}_l = \underline{U}_1$

$$\underline{I}_y = \underline{I}_l = \underline{I}_1$$

$$\underline{U}_1 = \underline{U}_2.\cosh\gamma.l + \underline{Z}.\underline{I}_2.\sinh\gamma.l \quad (3)$$

$$\underline{I}_1 = \underline{I}_2.\cosh\gamma.l + \frac{\underline{U}_2}{\underline{Z}}.\sinh\gamma.l \quad (4)$$

For

$$\underline{Z}_{20} = \infty, \quad \underline{I}_{20} = 0, \quad \underline{U}_2 = \underline{U}_{20} \quad (5)$$

$$\underline{U}_{10} = \underline{U}_{20}.\cosh\gamma.l = \underline{U}_{20}.\cosh j.\beta.l = \underline{U}_{20}.\cos\beta.l \quad (6)$$

$$\underline{I}_{10} = \frac{\underline{U}_{20}}{\underline{Z}}.\sinh\gamma.l = \frac{\underline{U}_{20}}{\underline{Z}}.\sinh j.\beta.l = j.\frac{\underline{U}_{20}}{\underline{Z}}.\sin\beta.l \quad (7)$$

$$\underline{Z}_{10} = \frac{\underline{U}_{10}}{\underline{I}_{10}} = \frac{\underline{U}_{20}.\cos\beta.l}{j.\frac{\underline{U}_{20}}{\underline{Z}}.\sin\beta.l} = -j.\underline{Z}.\mathrm{cotg}\left(\frac{2.\pi}{\lambda}.l\right)$$

where $\underline{Z} = \sqrt{L/C}$

Note: Variables $\mathcal{U}$, $\mathcal{I}$, $\underline{Z}$ are complex values; *U*, *I*, *Z* are their modules.

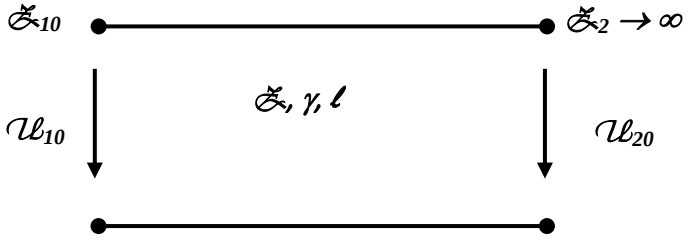


Figure 4. The branch line as a no-load-circuit line

From (7), it is emerged that for lengths $l = (2n+1).\lambda/4$, where $n = 0, 1, 2, \ldots$, the value $\underline{Z}_{input} = 0$, so the branch line causes the shortcut on the subscriber shunt. The subscriber line between xDSL modems should be flat (without branch lines) and must satisfied the standard norms. In the case that other equipment is connected to the subscriber line and satisfies above-mentioned conditions, it is necessary to adjust the length of the branch line so that $l \notin \langle (2n+1).\lambda/4 \pm \lambda/8 \rangle$ for $n = 0, 1, 2, \ldots$ For $f = 6$ MHz, $v_f = 2.10^8$ m.s^{-1}, $\lambda = v_f / f \approx 33{,}3$ m, the branch line length needs to be poles apart 8,3 m and odd multiples.

III. LIMITS OF THE CHANNEL CAPACITY FOR TRANSMISSION CHANNELS

Except negative environmental influences to those belong primarily influences of the white noise AWGN, the NEXT and FEXT crosstalks or their combination, the transmission channel capacity is limited mostly by internal parameters [11]. Features of the homogeneous symmetric line are given by basic material and geometric characteristics of the wires, which are expressed in the form of the primary constants (R_0, L_0, C_0, G_0). The solution of the telegraph equation system introduces the secondary constants $\underline{Z}$ and γ. In the case of perfectly terminated lines, it is possible to get the transfer function $\mathcal{H}(l, f)$ of the channel as

$$\mathrm{H}\ (l,f) = e^{-\gamma(f).l} \tag{8}$$

The importance of the channel capacity cognition is following: If the information rate R from the source is lower than the channel capacity C ($R < C$), then is theoretically possible to attain the reliable, errorless signal transport through the channel using appropriate coding techniques. On the other side, if the information rate is higher than the channel capacity ($R > C$), the errorless transport is impossible irrespective of an amount of the signal processing realized in the transceivers. This is the reason, why it is useful to know theoretical limits for given transmission path. In the environment of homogeneous symmetric lines, the following formula is valid for the calculation of the channel capacity C

$$C = \int_0^{\infty} \log_2\left(1 + SNR(l,f)\right) df = \tag{9}$$
$$= \int_0^{\infty} \log_2\left(1 + \frac{\left|\mathrm{H}\ (l,f)\right|^2 PSD_s(f)}{N(f)}\right) df \quad [\mathrm{bits/s}]$$

where $SNR\ (l, f)$ is the signal-to-noise ratio function in this environment, $PSD_s\ (f)$ is the power spectral density function of the information signal and $N\ (f)$ is the spectral function of the noise.

It is obvious that the alone core diameter of the copper wire has a significant influence on the determination of limiting condition of transmission channel capacity (Fig. 5). Fig. 5 displays the channel capacity C [Mbit/s] as a function of the line length l [km] for lines with various core diameters of wires at the input signal power 10 mW in the environment of the dominant white noise AWGN with the increased level -110 dBm/Hz. It can be seen that theoretical upper limits are significantly higher for wires with greater diameters. This advantage of wires with $\phi = 0{,}63$ mm and primarily $\phi = 0{,}9$ mm core diameters is much significant at longer line lengths and enables to predict another possibilities of the xDSL technologies development.

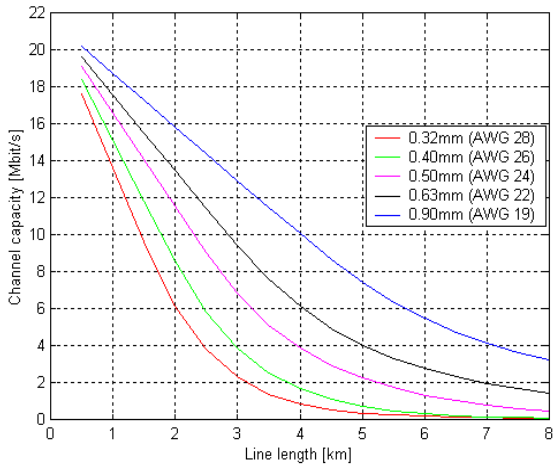


Figure 5. Channel capacity's characteristics versus line lengths for various core diameters of wires

While the channel capacity C determines limits of an error-free transmission at the infinite circuit complexity, the cut-off rate R_0^* is regarded as a practical transmission limit for the moderate coding complexity. The difference between the channel capacity and the cut-off rate nears its asymptotic value quite quickly increasing of the SNR value and the minimum cut-off rate R_0^* by calculating the channel capacity alone is destined by the following relation:

$$R_0^* \geq C - 0{,}2787\ [\mathrm{bits/symbol}] \tag{10}$$

Using these formulas, we calculated the maximum throughputs (theoretical and practical) possible to attain on metallic homogeneous lines [11].

From graphical characteristics of the channel capacity and cutoff rate, it is emerged that actual metallic lines are not effectively utilized for the information signal transport at very high-speed rates and that there still exists possibilities to reach a increasing of transmission rates closer to their practical limits for both downstream and upstream directions of the transport.

IV. Conclusion

For the successful extension and utilization of the multimedia services between current customers, it is necessary to solve the "bottleneck" problem in the access networks that are mostly compound from homogeneous symmetric metallic lines. The xDSL technologies allow the fast and comfortable solution of this problem. However, it is unavoidable to find out the real and actual state of subscriber lines before the exploitation of the xDSL technologies. In the case of non-standard parameters, it is imperative to realize such measurements and activities that remove the places of the line defects and impairments.

In the context with the analysis of theoretical limits of the transmission channel capacity, it is appropriate to consider about exchanges of older metallic lines with lesser core diameter wires for new homogeneous symmetric lines with greater core diameter wires (ϕ = 0,9 mm, ϕ = 0,63 mm). This allows ensuring the perspective utilization of common metallic access networks in the future, concretely at the utilization of the VDSL technology in the combination with the PON optical network.

At low values of the SNR ratio, the main environmental negative influence is the AWGN, because this type of the noise is always presented on the line, and the NEXT and FEXT noise powers are low due to low input signal powers. In this case, the difference between bandwidth efficiencies expressed in bit/s/Hz for the channel capacity C and cutoff rate R_0^* is increasing. However, at high values of input signal powers, the value of the SNR ratio is also increased and the NEXT crosstalk becomes a dominant influence. In this case, parts of the curves for the bandwidth efficiency for C and R_0^* are nearly parallel and the difference between them becomes unvaried.

ACKNOWLEDGMENT

This work is a part of research activities conducted at Slovak University of Technology Bratislava, Faculty of Electrical Engineering and Information Technology, Department of Telecommunications, within the scope of the project VEGA No. 1/0106/11 “Analysis and proposal for advanced optical access networks in the NGN converged infrastructure utilizing fixed transmission media for supporting multimedia services”.

REFERENCES

[1] J. Čuchran, R. Róka: The analysis of the feasibilities of the utilizing of the xDSL technologies in the access networks of Slovak Telecom. The final report of the technical project ZoD 1163, February 1999, Bratislava, Slovakia

[2] K.D. Kammeyer: Nachrichtenübertragung. B.G.Teubner, Stuttgart, 1992

[3] M. Vaculík: Access networks. ŽU EDIS Publishing house, Žilina, 2000

[4] P. Lafata, J. Vodrážka: Modeling of Transmission Functions and Crosstalk in Metallic Cables for Implementation of MIMO Concept. In: Radioengineering, Vol. 18, No. 4, 2009, pp. 491-496, ISSN 1210-2512

[5] J. Vodrážka, J. Sýkora: New Methods for Subscriber Line Modeling. In: 11th International Conference RTT 2009, CTU Prague, Czech Republic, pp. 67-70, ISBN 978-80-01-04410-0

[6] J. Vodrážka, J. Hrad: Modeling of a Subscriber Line with Inhomogenity. In: 32nd International Conference TSP 2009, Dunakiliti, Hungary, pp. 79-83, ISBN 978-963-06-7716-5

[7] R. Róka: Modeling of Environmental Influences at the Signal Transmission by means of the VDSL and PLC Technologies. In: International Journal of Electrical Communication Networks and Information Security – IJCNIS, Vol. 1, No. 1, 2009, pp. 6-13, ISSN 2073-607X

[8] R. Róka: Environmental Influences on the Power Spectral Densities of VDSL Signals. In: Journal of Electrical Engineering – EČ, Vol. 55, No. 1-2, 2004, pp. 18-24, ISSN 1335-3632

[9] J. Čuchran, R. Róka: The Internet service extension in the access networks by means of xDSL technologies. In: Telecommunications 2000, Bratislava, Slovak Republic, pp.182-185, ISBN 80-968042-7-8

[10] J. Čuchran: Telecommunication lines. STU Publishing house, Bratislava, Slovakia, 2008

[11] R. Róka: Theoretical and practical limits of transmission channels used by the ADSL technology in the metallic line’s environment in the access network in Slovakia, In: Journal of Electrical Engineering – EČ, Vol. 53, No. 3-4, 2002, pp. 91-96, ISSN 1335-3632

Routing Management Based on Statistical Cross-Layer QoS Information Regarding Link Status

Melinda Barabas, Georgeta Boanea, Andrei Bogdan Rus, Virgil Dobrota
Communications Department
Technical University of Cluj-Napoca
Cluj-Napoca, Romania
Email: {Melinda.Barabas, Georgeta.Boanea, Bogdan.Rus, Virgil.Dobrota}@com.utcluj.ro

***Abstract*—This paper presents the design principles and the practical implementation of a routing management solution which takes into account statistical cross-layer Quality of Service information regarding the state of the network. Link monitoring and communication between neighboring nodes is realized by the routing management system, while the routing protocol deals with routing tables and packet forwarding. We propose two types of management entities: *a)* local management entity (LME) responsible for monitoring directly connected incoming links and detecting traffic changes and *b)* domain management entity (DME) which captures the global state of the network and informs routers to avoid congested areas. The goal is to offer physical layer information to the routing algorithm, i.e. the metrics will depend on the traffic load and the delay measured on individual links. By identifying congested areas and anomalies, we can ensure a situation-aware path selection, improving the global capacity of the network by determining optimal routes.**

***Index Terms*—congestion control; cross-layer QoS; routing management; self-management; traffic-aware shortest paths.**

I. Introduction

Routing management is defined by different types of management functions necessary for the efficient operation of the routing process. A significant part of this consists of the monitoring procedures which help to understand what is going on inside the network. Traditional management relies on human supervision and monitoring, as well as manual interventions, in order to ensure that the system operates as desired. Traffic can be manually rerouted, but only after an extended service outage. But new networks and services are becoming more complex and thus new approaches are needed. The authors of [1] summarize the state of worldwide research in the domain of management and propose design requirements for Future Internet management solutions. New trends focus on increasing the level of automation, developing scalable and intelligent self-management systems which do not require human intervention, as presented in [2].

The role of management for the different routing mechanisms is discussed in [3]. Initially, it was believed that routing management was not necessary because the adaptive routing protocols, such as OSPF (Open Shortest Path First), can react to failures and changes in the network. But efficient traffic engineering is not possible if routing protocols do not take into consideration the real-time traffic conditions.

In the case of OSPF, the most widely used routing protocol in large networks, link connectivity is verified via Hello messages sent every 10 seconds. If no Hello message is received from a neighbor for 40 seconds, it is marked as being down. Thus, OSPF is not able to solve the problem of congestion detection. If Hello messages are missing because of congestion, apart from the fact that in the majority of cases congestion is unidirectional, OSPF eliminates the possibility of using the affected link in both directions. In addition, all links through that router are also marked as down and will not be used. Thereby, the initial traffic routed through this node will be transmitted on other links, reducing the load balancing in the network and increasing the possibility that other congested areas appear. Another problem is the fact that if a link flaps constantly due to congestion, but at least 1 out of every 4 Hello messages is received, OSPF does not detect the cause of missing packets.

To solve the problems mentioned above, a routing management system is proposed which offers cross-layer QoS information to the routing algorithm, helping it to avoid congested links and determine optimal routes.

The remainder of the paper is organized as follows. In Section 2 the main objectives of a routing management system are identified and the characteristics of the proposed system are described. Section 3 presents the designed management entities and the interaction between them. In Section 4 the practical testbed used for performance evaluation is described, followed by the experimental results in Section 5. Finally, Section 6 concludes the paper and discusses future work.

II. Routing Management System

QoS-aware routing has become a great challenge in computer networks. It refers to guaranteeing traffic delivery in terms of bandwidth, delay and packet loss. Traditional QoS models are presented in [4]. Usually, QoS-aware routing algorithms, like the one described in [5], assume source routing and allocate bandwidth in advance, according to traffic parameters. Besides resource admission control, traffic shaping and

dropping can also be used in order to meet QoS requirements. These solutions have some disadvantages, for example the waste of bandwidth on the entire path reserved in advance or packet losses due to special packet drop policies.

In this paper we present the design principles and the practical implementation of a distributed routing management system. The main objectives of the approach described herein are the following: *a)* real-time monitoring of network parameters; *b)* detecting events, anomalies and congestion in order to estimate the state of the network; *c)* communicating with the routing protocol in order to recalculate the routes and *d)* self-management. Through these functionalities QoS-awareness can be achieved.

The main characteristic of our routing solution consists of the idea of separating the link monitoring and the update of network state information from the routing itself. The routing management system takes over the monitoring functionalities and all the communication between neighboring nodes, while the routing protocol deals with the routing decisions and packet forwarding. Hence, the information gathered by the routing management system regarding the state of the network can be reused for other purposes.

The routing management system is composed of management entities which collect measurement data through monitoring. Link status information is gathered through cross-layering and is used for QoS-aware routing. The goal is to determine the traffic load and the delay on each link in order to offer an overview of current network resource availability. Based on the topology and the network state, the routing algorithm running on a given node determines the routing table of that node. Interaction between neighboring nodes is carried out just by the routing management system; no other routing messages are exchanged. The proposed routing management solution represents a situation-aware self-managing system which is capable to dynamically adapt to external events, minimizing the need for human intervention.

The solution presented in this paper relies on cross-layering techniques which allow communication between non-neighboring layers of the OSI model. The implemented cross-layer QoS mechanism is described in detail in [6] and [7]. Information derived from the physical layer is passed to the upper layers to assist the routing management system in determining the operation modes of routing protocols. Two types of QoS parameters are used, reflecting link status at the Physical Layer and MAC Sub-layer: *a) available transfer rate* (ATR) which is obtained through passive monitoring and *b) one-way delay* (OWD) which is measured in real-time for each direction of each link through an active monitoring approach (i.e. by injecting probes). OWD consists of processing-, queuing-, transmission- and propagation delay. When routers receive the link state information (ATR and OWD), they can compute the traffic-aware shortest paths for optimal network performance.

III. Designing the Management Entities

Two types of management entities were defined: *Local Management Entity* (LME) and *Domain Management Entity* (DME). In this section we describe the functionalities and the operation mode of the implemented entities, as well as the interaction between them.

Our solution is a software application running on Linux-based machines. Management, measurement and routing applications are written in C++ and Qt under Fedora operating system, using MySQL database with mysql++ connector. Management entities located on different nodes communicate through XML messages, storing the network status information in local databases.

A. Local Management Entity

Local Management Entities are software applications located on every node of the administered system. These capture the local network state by observing, collecting and interpreting data and by communicating with neighboring nodes.

The main functionalities of such an entity are the following:

- *Starting the cross-layer measurement applications for monitoring ATR and OWD on the directly connected inbound links.* Resource monitoring is made periodically, second by second.
- *Using statistical indicators to detect changes on individual links.* These can indicate not just the real-time values, but also the trend of the evolution for a specific parameter. At the moment we are using two different indicators: *simple moving average* (SMA) calculated every second for an interval of $N = 10$ seconds in order to estimate the value of the monitored parameter, and *standard deviation* (σ) for identifying events/changes of the traffic conditions.

$$SMA(t) = \frac{1}{N}\sum_{i=0}^{N-1} x(t-i)\ . \tag{1}$$

$$\sigma(t) = \sqrt{\frac{1}{N}\sum_{i=0}^{N-1}\left(x(t-i) - SMA(t)\right)^2}\ . \tag{2}$$

- *Extracting the current status of the local network from the collected cross-layer QoS information.* The local network refers to the directly connected incoming links.
- *Communicating with other LMEs and DME over UDP or TCP connections by sending XML messages.*

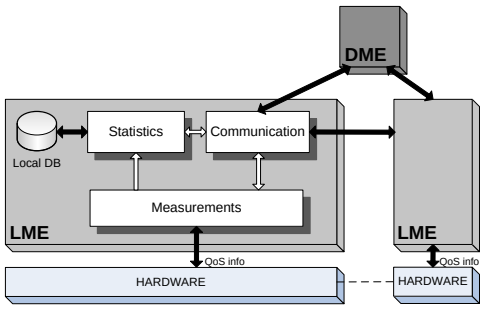


Fig. 1. Local Management Entity

Fig. 1 shows the main building blocks of the LME. The *measurement module* manages the monitoring applications.

It receives the ATR and OWD measurements through cross-layering techniques and makes these values available to the statistics block. The *statistics module* calculates the SMA and the standard variation of the monitored network parameters and identifies changes in the link status. This module is also responsible for storing the status information in a local database. The *communication block* is responsible for transmitting messages to other entities.

The flowchart diagram in Fig. 2 depicts the operation of such an entity. LME starts monitoring the local links as it detects the corresponding neighboring nodes. The SMA and the standard deviation are calculated for a certain time period, this is done every second. The statistical indicators are compared with predefined threshold values ($th1$ and $th2$). If $SMA < th1$ or $\sigma > th2$, the DME is announced that an event occurred. Because $N = 10$ in equation (1), congestion can be detected after at least 10 seconds. We chose this value as a compromise between the possibility of generating false alarms and the reaction time of the system.

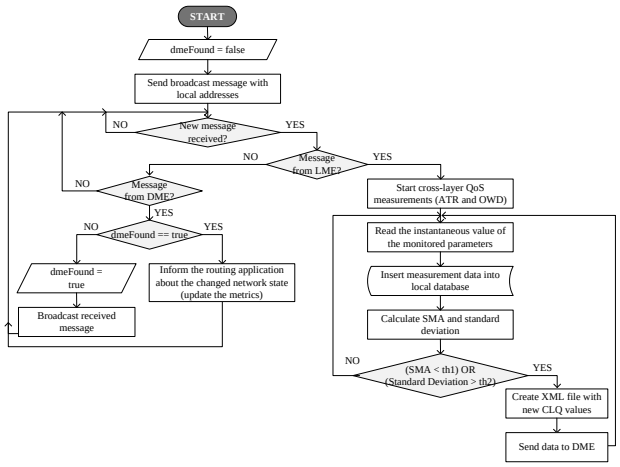


Fig. 2. Operation mode of LME

B. Domain Management Entity

To address the problem of scalability, we consider that the network is divided into domains. A domain management entity corresponds to a domain (i.e. a group of routers), and is dealing with problems that require a global vision of the network (e.g. routing metrics). In the case of routing management, the system has to provide the same information regarding the network state in all nodes. Otherwise the routes could not be determined correctly. This means that we have to centralize this information and send it simultaneously to the routers.

The modules of the DME are illustrated in Fig. 3. Through the *communication module*, the DME interacts with LMEs. The *statistics module* is responsible for storing the status information in a local database. The domain management entity was designed to have a module for *network state prediction*, but this is not yet implemented practically. We intend to use artificial neural networks (NN) for time series forecasting, allowing us to detect congestions before they appear. Thus,

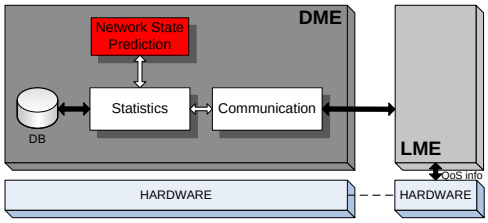


Fig. 3. Domain Management Entity

the routing management system would be able to take actions in order to avoid these affected areas, reducing packet losses.

The DME can be located on any of the nodes of the domain. The existence of such an entity is not just motivated by the need to centralize data, but by performing functions exceeding the scope of LMEs. It is able to perform more complex operations that require additional resources, such as prediction. We chose this approach because forecasting is a resource consuming task in the means of memory and processing power. Most nodes in a network have limited capabilities, thus adding prediction functionalities in every single node is not feasible.

C. Interaction between Entities

As mentioned before, different management entities interact by sending XML messages. In Fig. 4 we can see the communication between the different types of entities. This represents a very simple scenario with the purpose of illustrating the message-passing between entities.

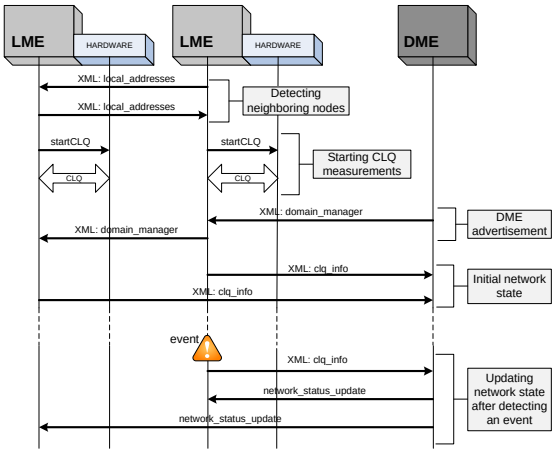


Fig. 4. Interaction between LMEs and DME

We can identify five major steps in the presented message flow:

1) *Detecting neighboring nodes:* The LMEs periodically advertise their local addresses in order to detect directly connected nodes (this also includes topology discovery) and to start monitoring the corresponding links.
2) *Starting the cross-layer QoS measurements:* After detecting the direct neighbors, the measurement applications are started. Each LME will monitor the ATR and the OWD on incoming links. Passive monitoring is

used for ATR, and active monitoring for OWD (i.e. by sending probes to the neighboring nodes).

3) *Advertising the presence of DME:* DME periodically advertises its presence by sending XML messages which contain its IP address and the destination port.
4) *Determining initial network state:* When LMEs detect the presence of the DME, they will send an XML message with initial link status information. In this way, the domain manager can have an initial global view of the network topology and the status of the network.
5) *Updating network state after event detection:* The network status updates are event-driven. Through the congestion discovery mechanism, a LME detects changes in traffic conditions on a monitored link and informs the DME. DME sends the global state of the network to LMEs at maximum 2 hop distance from the affected link, limiting the amount of signaling and the number of routers which will have to make changes to the routing table. TCP connections are used for this type of interaction in order to ensure that the messages are received.

Routers will update their routing tables according to actual traffic conditions indicated by DME, selecting the route with the largest ATR and the lowest delay. When selecting the path, the routing protocol uses the idea of the Ford-Fulkerson algorithm, i.e. as long as there is no congestion, the current paths are not changed. By using this technique, we avoid frequent oscillation of the routing process, which could cause out-of-order packet arrivals and performance degradation.

But if a link is found to be congested, not all traffic will be rerouted because otherwise the link would remain unused. At a given moment only loss-sensitive streams (e.g. video, audio) belonging to a *source–destination* pair will be transfered to a different path. For this, traffic flows affected by packet losses and delays have to be identified. The network state is reevaluated and if the congestion persists other stream will be also re-routed. This is possible because we foresee a *multipath routing solution* to be used. This mechanism allows for balancing the traffic load sent between a source–destination pair on different links.

IV. Performance Evaluation

For performance evaluation we selected the test architecture presented in Fig. 5. We chose this network topology because it offers sufficient paths between the source and destination nodes, but at the same time it is simple enough to allow practical implementation using physical machines. Beside the source and destination, the testbed contains six routers (R1, R2, R3, R4, R5 and R6) running on Linux-based machines with Fedora operating system. The computers have Intel Core 2 Duo CPUs running at 2.66 GHz and 4GB of memory. On each machine runs a QoS-aware multipath routing application and the LME application. In addition, on R5 there is a domain management entity running.

The experimental parameters are shown in Table I.

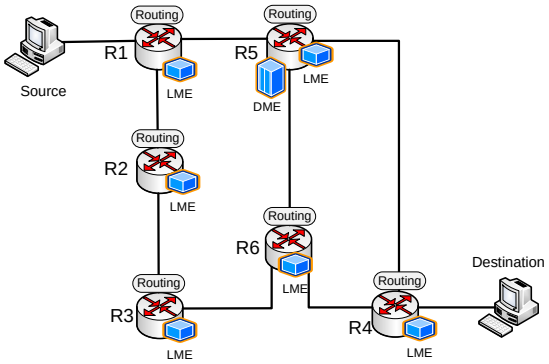


Fig. 5. Practical testbed

TABLE I
Experimental parameters

Parameter	Value
Number of routers	6
Background traffic generator	iperf
Multimedia traffic generator	VLC
Experiment duration	3 minutes
Congestion appearance	after 1 minute

To demonstrate the beneficial effects of cross-layer QoS-aware routing, we chose to use video streams in the experiments because the Internet is experiencing a substantial growth of video traffic. Providing good video quality is a major problem since video traffic is both massive and intolerant to packet loss or latency. According to Cisco VNI (Visual Network Index) forecasts, the sum of all forms of video will exceed 91% of global consumer traffic by 2014 [8].

For the test scenario, we assume that MPEG-4 VBR (Variable Bit Rate) video flow is sent over RTP/UDP/IP between the source-destination pair. The video stream has an average bitrate of 1.5 Mbps and it is transmitted for a period of three minutes, by a VLC client running on the source. The IP packets sent by the VLC client contain 1356 bytes (RTP/UDP/IP headers: 40 bytes; multimedia payload: 1316 bytes). The main parameters of the video stream are presented in Table II.

TABLE II
Video stream parameters

	Parameter	Value
Encapsulation		MP4
Video	Codec	H.264
	Resolution	624×352
	Frame rate	25
Audio	Codec	AAC
	Sample rate	48 kHz
	Channels	Stereo

To introduce congestion between R5 and R4, we used `ethtool` to limit the nominal rate of the link to 100 Mbps (the actually observed transfer rate at the network layer is less than this theoretical maximum). After this, we generated background traffic between R5 and R4 using `iperf`: 10 UDP streams, with the total transfer rate of 100 Mbps.

The goal of the tests carried out is to demonstrate the capability of the routing management system to reduce the negative effects of congested links by providing statistical cross-layer QoS information regarding link status to the routing application. We performed experiments in order to compare the behavior of the following routing approaches in case of congestion on one or more links: *Case 1)* OSPF (Open Shortest Path First) and *Case 2)* QoS-aware routing application communicating with the routing management system.

The evaluation of the performance of each routing solution is based on the parameters of the received video stream. We take into consideration the number of lost packets, determined by examining the Sequence Number field in the RTP (Real-time Transport Protocol) header and observing if there is a gap in the counter. Besides the percentage of lost packets, other two video quality metrics are also used for performance evaluation, namely: *a)* magnitude of loss events and *b)* discontinuity counter. The magnitude of loss events expresses the number of packets that were dropped at each loss event (note that a magnitude of 0 means the packet arrived successfully at the destination). The discontinuity counter measures discontinuity events, characterizing the frequency with which discontinuities were detected, i.e. the number of times loss was detected.

V. Experimental Results

If there is no congestion in the network, both approaches use the same path for routing the packages (source–R1–R5–R4–destination) and no packets are lost during the streaming.

In the experiments, we suppose that after 1 minute of streaming, link R5–R4 starts to be affected by congestion due to background traffic generated with the `iperf` network testing tool. As a result, ATR drops below the required rate to transmit the stream and the OWD from node R5 to R4 increases.

Case 1:

Fig. 6 illustrates the testbed for *Case 1*. After we introduce congestion between R5 and R4, OSPF does not modify the path between the two nodes because it does not take into consideration the physical state of the links. As an effect, we observe packet losses at the destination node, the quality of experience being very poor. In this case 6461 packets where missing at the destination out of 21980 transmitted packets, i.e. 29.39% of the transmitted packets was lost.

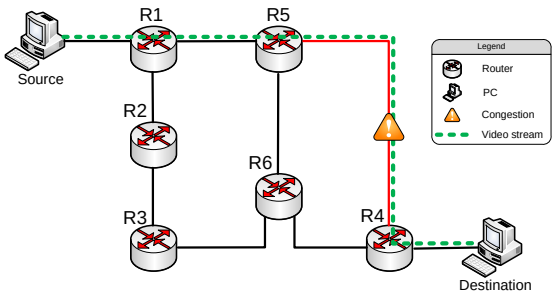


Fig. 6. Case 1 – Testbed

In Fig. 7 the magnitude of loss events is presented. The average magnitude of loss events is 0.416 and the maximum value is 16. The discontinuity counter is equal to 4413.

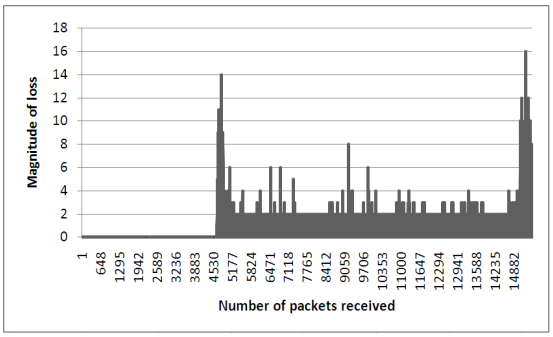


Fig. 7. Case 1 – Magnitude of loss events

Case 2:

In this case, the congestion is detected by the local management entity located on router R4. The QoS-aware routing solution is notified by the routing management application about the congestion and it determines the new route from source to destination node, based on the updated routing metrics. In this scenario, presented in Fig. 8, the new path which is written into the routing tables will be source–R1–R5–R6–R4–destination. The quality of the streaming is affected just for a short period of time, as a result of congestion detection and router reconfigurations. The percentage of lost packets was 3.37%; this means that only 741 packets got lost out of 21980.

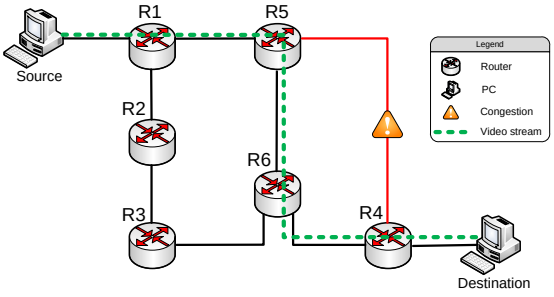


Fig. 8. Case 2 – Testbed

Table III depicts the number of packets sent and received in the two cases analyzed earlier.

TABLE III
Comparison between the two cases

	Case 1	Case 2
Sent packets	21980	21980
Received packets	15519	21239
Lost packets	6461	741
Percentage of lost packets	29.39%	3.37%
Average magnitude of loss	0.416	0.035
Maximum magnitude of loss	16	11
Discontinuity counter	4413	273

Fig. 9 illustrates the magnitude of loss event in *Case 2*. It can be observed that losses appear just for a short period of time, until the stream is rerouted. In this case, the average magnitude of loss events is 0.035, with a maximum value of 11. The discontinuity counter is 741.

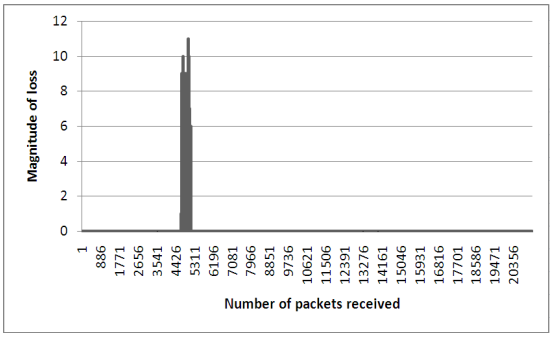


Fig. 9. Case 2 – Magnitude of loss events

In Fig. 10 the evolution of the success ratio is shown over the experiment duration. The success ratio at a given moment is determined by the ratio of the number of packets received successfully and the total number of packets sent. In *Case 1* the success ratio is falling steadily upon the appearance of congestion, reaching a minimum of 70.61% at the end of the experiment. In *Case 2* the success ratio drops for a short time to 87.4%, but after that it recovers, increasing to the final value of 96.63%.

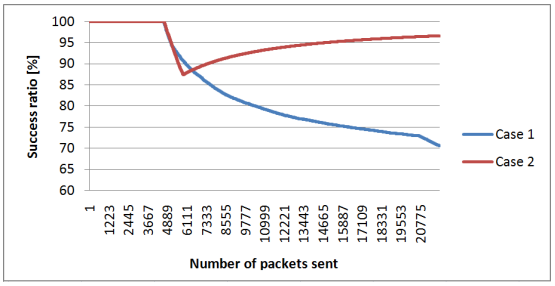


Fig. 10. Success ratio over experiment duration

The percentage of lost packets is 29.39% when using OSPF, whereas in the case of our solution it drops to 3.37%. As a conclusion, we can say that due to routing management a new route can be selected which avoids the congested link and ensures the lowest OWD. Thereby, the percentage of lost packets can be reduced significantly which results in a better quality of experience. Thus, the routing management system can complete the routing algorithm, offering a more reliable media streaming. Not only the media streaming benefits from this rerouting, but also the other transmissions on the affected link because in this way the congestion can be alleviated or can disappear. This comes with the cost of introducing additional traffic overhead necessary for transmitting management information throughout the network.

Regarding the average hardware resources, the LME application uses 0.3% of the CPU and 0.1% of the memory of the PC, while the DME application uses 0.8% of the CPU and 0.1% of the memory.

VI. Conclusions and Future Work

Summarizing the presented routing management system, we can tell that this is a self-managing system. This means that after setting the management policies in the configuration files and starting the applications on every node, no further human interaction is needed.

The proposed routing management solution takes into account statistical cross-layer QoS information regarding the state of the network. These are distributed throughout the network and are used by the routing protocol as routing metrics. Thus, the global capacity of the network can be improved by avoiding congested areas. The idea was to separate the monitoring functions from the actual routing, making the statistical network status information reusable. The routing protocols can focus on path determination and updating the routing tables.

This is a preliminary implementation, covering only a part of the possible functionalities of a complete routing management system and offering a proof-of-concept. As future work, we intend to evaluate the effect of prediction in terms of the compromise between the utilized resources and the improved global capacity of the network. We envisage reducing the packet losses through congestion prediction as opposed to congestion detection used in the present implementation.

Acknowledgment

This work was partially funded by the PRODOC program ”Project of Doctoral Studies Development in Advanced Technologies” (POSDRU/6/1.5/S/5 ID 7676).

References

[1] Sung-Su Kim, Y. J. Won, , Mi-Jung Choi, J. W.-K. Hong, J. Strassner, *Towards management of the Future Internet*, 11th IFIP/IEEE International Symposium on Integrated Network Management. Long Island, USA, 2009.

[2] R. Roth, F. Wolff, T. Zseby (eds.), *D-4.1 Definition of scenarios and use cases*. FP7-ICT-2007-1-216041-4WARD – Architecture and Design for the Future Internet, 2009.

[3] D. Medhi, *Routing Management in the PSTN and the Internet: A Historical Perspective*, Journal of Network and Systems Management, Volume 15, pp. 503–523. Springer, 2007.

[4] M. S. H. Lipu, M. N. -U. Haque, T. F. Karim, M. L. Rahman, M. H. Morshed, , *Quality of service model selection criteria and traffic controlling algorithm of a native IP Network*, International Conference on Educational and Information Technology ICEIT 2010, pp. V3-144 – V3-148. Chongqing, China, 2010.

[5] E. Marin-Tordera, X. Masip-Bruin, S. Sanchez-Lopez, J. Domingo-Pascual, A. Orda, *The Prediction Approach in QoS Routing*, IEEE International Conference on Communications ICC 2006, pp. 1020–1025. Istanbul, Turkey, 2006.

[6] A. B. Rus, M. Barabas, G. Boanea, Zs. Kiss, Zs. Polgar, V. Dobrota, *Cross-Layer QoS and Its Application in Congestion Control*, 17th IEEE Workshop on Local and Metropolitan Area Networks LANMAN 2010. Long Branch, USA, 2010.

[7] L. M. Correia, H. Abramowicz, M. Johnsson, K. Wuenstel (editors), *CLQ-Based Testbed used for Generic Path*, Chapter 12 ”Prototype Implementations”, pp. 271-276, Architecture and Design for the Future Internet. 4WARD Project. Series: Signals and Communication Technology. 1st Edition, Springer Science + Business Media, 2011.

[8] *Cisco Visual Networking Index: Forecast and Methodology, 2009–2014*, White Paper, 2010.

Optimization of the NGOSS Change Management Process Simulation Model

Michal Mrajca, Jakub Serafin, Zdenek Brabec
Department of Telecommunication Engineering, FEE, Czech Technical University in Prague
Prague, Czech Republic
mrajcmic@fel.cvut.cz

Abstract— **It is necessary for any telecommunication provider to introduce new products or changes of existing products as quickly and as often as possible. It helps them to stay ahead of competition. To achieve that they have to have an effective and efficient inner business processes. Almost every conscious company measure, assess and adjust its processes. This paper describes the Change Management process within New Generation Operation Support System (NGOSS) of one Czech telecommunication service provider, introduces its optimization with obtained results.**

Keywords—Change Management, NGOSS, probability, process modeling, process simulation, process optimization

I. INTRODUCTION

The Change Management process is still more important due to frequent changes in provided products and services. It becomes a vital part of a process portfolio of any service provider. Its efficiency and effectiveness is necessary to stay competitive in the dynamic market [1]. The Change Management process (as any other process) can be described as a single machine with the specific arrival distribution of the Request for Changes (RFC). By an arrival of the RFC it is meant a registration of the necessary change into the Change Management process. The processing of the RFC by a single machine is usually described by a probabilistic distribution function or a density function which specifies time it takes from registration the RFC till release of the change. It is common that RFCs arrive much faster than they are processed by the single machine. This fact causes that we need to introduce a buffer. The buffer can't be usually easily defined as first in first out (FIFO) because the RFCs have at least several priority levels – the process is more effective and efficient due to this fact [2]. The Change Management process described by a single machine process is illustrated on figure 1.

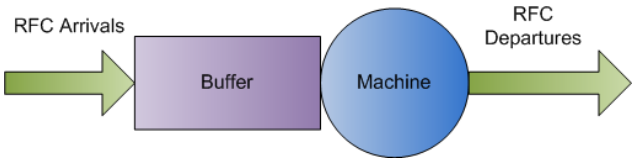


Fig. 1. Single machine [3].

To be able to better monitor, measure and manage the Change Management process we had introduced several milestones into this process and assessed the times the process

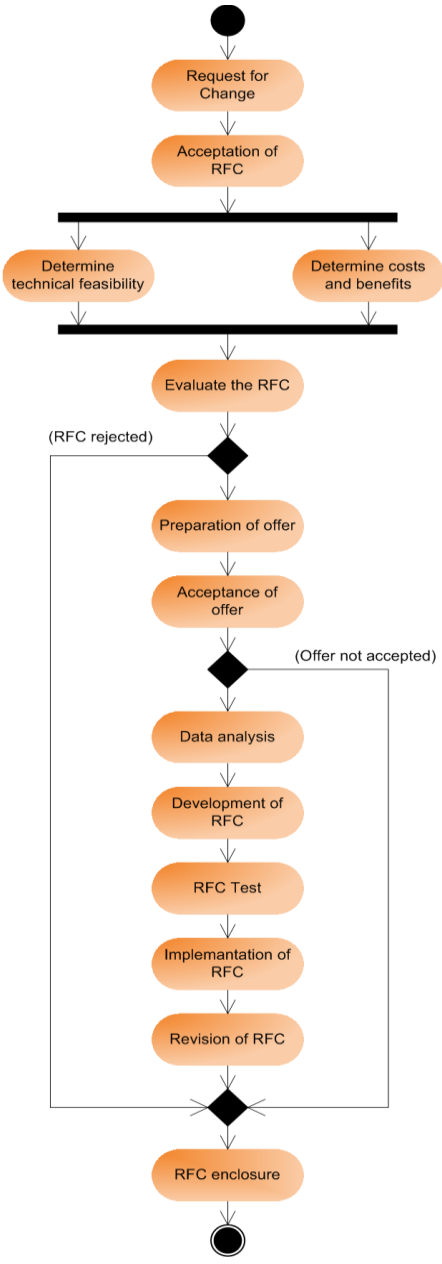


Fig. 2. Change Management process meta-model.

needs to reach each milestone. Each milestone defines the end of appropriate activity (process step) of the process. The milestones weren't placed randomly. Their positions result from the common Change Management meta-model [2] which was slightly adjusted to the conditions of the modeled process. Final meta-model is illustrated on figure 2.

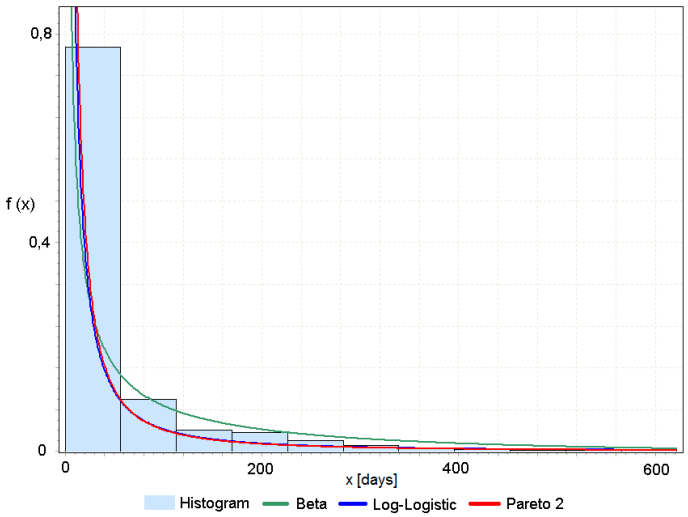


Fig. 3. Probability density function of "Evaluate the RFC" activity.

TABLE I
DISTRIBUTION FITTING FOR "EVALUATE THE RFC" ACTIVITY

Distribution	Kolmogorov - Smirnov		Anderson - Darling		Chi-Squared	
	Statistic	Rank	Statistic	Rank	Statistic	Rank
Pareto 2	0.18619	1	139.32	3	546.82	5
Beta	0.19196	2	141.82	4	524.03	4
Wakeby	0.20681	3	119.4	1	416.1	2
Log-Logistic	0.20681	4	119.4	2	416.1	3
Phased Bi-Exponential	0.27583	5	571.26	15	1946.4	24

To be able to monitor and asses times for each activity of the process, each change manager has to input the date and time of the end of the activity as well as information whether the process continues to next activity or process ends for any possible reason. We suppose that the next activity starts immediately after the end of the previous activity is saved into the system. The activity is bounded by two consecutive milestones.

The duration is then calculated in very simple way. For example the duration of the activity "Implementation of RFC" is calculated as difference between total process time at the end of activity "Implementation of RFC" and the time at the end of activity "RFC test" (see fig. 2). One RFC is then represented as a two dimensional array in the monitoring system.

Process can be described as one single machine (fig. 1). When we have an array of times we can describe it as a row of cooperative single machines. It allows as to build more precise model for simulation. For modeling purposes there weren't impeached distributions of the RFCs arrivals and type of buffers (see fig. 1). The research is strictly focused on a probability distribution of time that is needed for processing of the RFC between each pair of the milestones.

II. VALUES EVALUATION

The Change Management process was monitored and measured for more than three years. During this time arrived over two thousands RFCs into the monitoring system. 93% of the RFCs were processes, 47% of them were implemented.

For the evaluation of a time probability distribution between each pair of the milestones there were used three "goodness of fit" tests. These tests compare a random sample of data with a theoretical probability distribution function. In other words, these tests show how well the distribution we have selected fits to selected data [4].

Each activity has its own set of random time values. Each value specifies the time that RFC spent in the activity of the process till it meets all requirements allowing it to pass to following activity. These values were used in EasyFit for "goodness of fit" tests. All three performed tests are described in following paragraphs.

A. *Kolmogorov–Smirnov test*

The Kolmogorov-Smirnov test is used to decide if a sample comes from a population with a specific distribution. For a random variable X and a sample $\{x_1, x_2, ..., x_n\}$ the empirical distribution function X is defined as [5]:

$$F_X(x) = \frac{1}{n}\sum_{i=1}^{n} I(x_i \le x) \qquad (1)$$

where I (condition) is the indicator function, i.e., 1 if the condition is true and otherwise 0. The Kolmogorov-Smirnov test statistic is defined as:

$$D_n = \sup_x |F_x(x) - F(x)| \qquad (2)$$

where sup is the supremum of the set and F(x) is the theoretical cumulative distribution of the distribution being tested which must be a continuous distribution and it must be fully specified (location, scale, and shape parameters cannot be estimated from the data).

B. *Anderson-Darling test*

The Anderson–Darling test uses the fact that, when given a hypothesized underlying distribution and assuming the data does arise from this distribution, the data can be transformed to a uniform distribution. The transformed sample data can be then tested for uniformity with a distance test. The Anderson-Darling statistic (A^2) is defined as [5]:

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)\cdot[\ln F(X_i) + \ln(1 - F(X_{n-i+1}))] \qquad (3)$$

The hypothesis regarding the probability distribution is rejected in case that test statistic (A^2) is greater than the critical value obtained from a table.

C. Chi-Squared test

For the chi-square goodness-of-fit computation, the data are divided into k bins and the test statistic is defined as [5]:

$$\chi^2 = \sum_{i=1}^{k} (O_i - E_i)^2 / E_i \qquad (4)$$

where O_i is the observed frequency for bin *i* and E_i is the expected frequency for bin *i*. The expected frequency is calculated by:

$$E_i = N(F(Y_u) - F(Y_l)) \qquad (5)$$

where F is the cumulative distribution function for the distribution being tested. Y_u is the upper limit for class *i*, Y_l is the lower limit for class *i* and N is the sample size.

This test is sensitive to the choice of the bins. There is no optimal choice for the bin width [4]. The best choices should produce similar but not identical results.

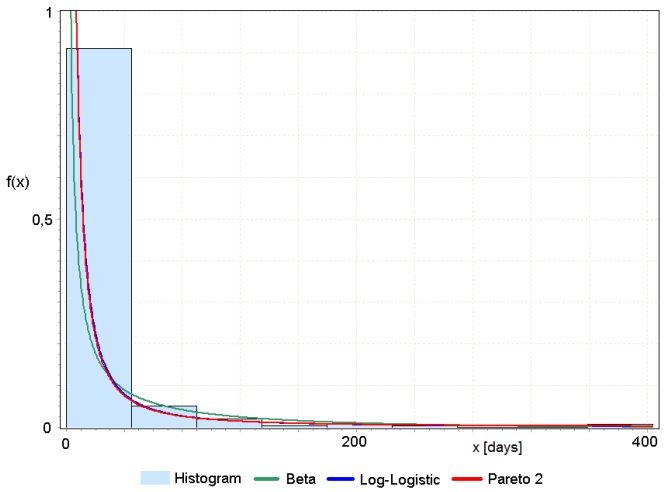


Fig. 4. Probability density function of "Data analysis" activity.

TABLE II
DISTRIBUTION FITTING FOR "DATA ANALYSIS" ACTIVITY

Distribution	Kolmogorov - Smirnov		Anderson - Darling		Chi-Squared	
	Statistic	Rank	Statistic	Rank	Statistic	Rank
Pareto 2	0.18619	1	6.1875	3	30.119	4
Beta	0.19196	2	5.1822	1	29.861	2
Gen. Logistic	0.20681	3	6.3821	4	57.462	5
Gen. Pareto	0.20681	4	66.509	13	120.42	12
Log-Logistic	0.27583	5	37.4	5	78.833	7

III. VALUES EVALUATION RESULTS

The measured values (a sample of random data) were compared with almost fifty distributions from Normal, Students or Longnormal to Beta or Erlang. The distribution fitting was supported by the distribution fitting software EasyFit. The fitting was done for every measured value between each defined milestones.

The results of the experiment are very interesting. For all steps there were almost the same distributions on the fore positions for all activities. The figures 3 and 4 illustrate probability density functions of two different activities ("Evaluate the RFC" and "Data analysis") of Change Management process. Tables I and II contain the distributions which fitted the best for these sets of values.

The parameters of the illustrated density functions of activities "Evaluate the RFC" and "Data analysis" are listed in the table III.

TABLE III
PARAMETERS OF ILLUSTRATED DISTRIBUTIONS

Process activity	Distribution	Parameters
Evaluate the RFC	Beta	α = 0.24059; β = 3.1991; a = 8.0529e-15; b = 1208.6
	Log-Logistic	α = 0.65139; β = 5.6125
	Pareto 2	α = 0.60063; β = 3.2559
Data analysis	Beta	α = 0.10474; β = 2.4346; a = 1e-14; b = 404
	Log-Logistic	α = 0.75142; β = 2.3445
	Pareto 2	α = 0.61199; β = 1.1742

The distributions which were assessed as the best fit for all activities of the Change Management process are Pareto 2, Log-Logistic and Beta. They are briefly introduced in the following paragraphs.

A. Pareto 2 distribution

The Pareto 2 or Lomax distribution is a skewed, heavy-tailed distribution that is sometimes used to model the distribution of income [6]. As with many other distributions, the Pareto distribution is often generalized by adding a scale parameter. Thus, suppose that *Z* has the basic Pareto distribution with shape parameter *a*. If $b > 0$, the random variable $X = b, Z$ has the Pareto distribution with the shape parameter *a* and the scale parameter *b*. Note that *X* takes values in the interval $[b, \infty]$. The Pareto 2 distribution is defined as [6]:

$$f(x) = \frac{\alpha \beta^{\alpha}}{(x+\beta)^{\alpha+1}}, x \geq \beta \qquad (6)$$

B. Log-Logistic distribution

The log-logistic distribution is the probability distribution of a random variable whose logarithm has a logistic

distribution. It is similar to the log-normal distribution in a shape but it has heavier tails. Its cumulative distribution function can be written in a closed form, unlike that of the log-normal. The parameter $\alpha > 0$ is a scale parameter and is also the median of the distribution. The parameter $\beta > 0$ is a shape parameter. The probability density function is defined as [7]:

$$f(x) = \frac{(\beta/\alpha)(x/\alpha)^{\beta+1}}{[1+(x/\alpha)^{\beta}]^2} \tag{7}$$

C. Beta distribution

The Beta random variable X assumes values in the unit interval [0, 1]. This interval can be scaled and shifted to any interval. The Beta distribution is denoted by Beta(α, β), where $\alpha > 0$ and $\beta > 0$ are two shape parameters and *a* and *b* and $a < b$ are scale parameters. The Beta random variable is often used in statistics to model an unknown probability and is defined as [7]:

$$f(x) = \frac{(x-a)^{\alpha-1}(b-x)^{\beta-1}}{B(\alpha,\beta)\cdot(b-a)^{\alpha+\beta-1}} \tag{8}$$

where:

$$B(\alpha,\beta) = \int_0^{\infty} y^{\alpha-1}(1-y)^{\beta-1}dy' = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \tag{9}$$

IV. Probabilistic Description of Process

Using the fact that the times of all activities of this process can be described by a few distribution functions, the NGOSS Change Management process can be modeled easily. The differentness of the distribution in each step of the process is so slight that there is possibility to use only one of the formerly mentioned distributions to model all activities of the NGOSS Change Management process.

To show the difference between models based on above mentioned results and model which uses only the best fit probabilities for each activity we have built several simulation models. We compared following models:

- All activities use the Pareto 2 distribution.
- All activities use the Log-logistic distribution.
- All activities use Beta distribution.
- All activities use their best fit (standard model).

The models were built in Arena simulation software which is based on SIMAN modeling language. During the simulation we observed three key performance indicators (KPI) of the process. They are:

- time of processing of each RFC,
- usage of resources,
- number of processed and rejected RFCs.

Third KPI is not affected by change of the time probability distribution. We focused on first two mentioned KPIs in the simulation results.

V. Results of Simulation

The simulation was focused only on the KPIs which can be affected by change of time probability distribution of the activity, namely time of processing of each RFC and usage of resources. The duration of simulation was set to one year. All results are shown in tables IV, V.

TABLE IV
RFC PROCESSING TIMES

Simulation model based on	Average process time [days]	Minimum process time [days]	Maximum process time [days]
Pareto 2	150.3570	6.3523	829.2547
Log-logistic	150.9841	7.1221	833.4873
Beta	148.6497	6.8143	837.4876
Best fit	149.6549	6.6709	832.3660

TABLE V
RESOURCE UTILIZATION

Simulation model based on	Change Manager Utilization [%]	Analyst Utilization [%]	Programmer Utilization [%]	Tester Utilization [%]
Pareto 2	68.14	83.78	91.47	81.97
Log-logistic	68.59	84.97	91.89	81.99
Beta	67.89	83.83	90.91	82.13
Best fit	68.35	84.24	91.24	81.45

The results show, that in case we select and use any of the proposed probability distributions for description of the time distribution of each activity, the difference of time the RFC spent in whole process is less then 0.87% for average time, 0.62% for maximum time and 6.76% for minimum time. Difference in resources utilization is less then 0.87% in comparison with usage of the best fit for each process step.

These results help us to build mathematically much easier model and to achieve shorter simulation times. The model which uses only one probability distribution for each process step is much easier for management – it is flexible and introduces changes easily.

VI. Conclusion and Future Work

The paper introduces results of the project which measured, modeled and simulated the NGOSS Change Management process of one of the Czech telecommunication service providers. The model of presented process can be described as a row of activities which have similar time distribution. This helps us to build much simpler model which is based on one of previously described probability distributions (Pareto 2, Log-Logistic or Beta). By usage of one probability distribution for all activities in the model it is much simpler to implement

changes of the process. On the other hand the usage of only one distribution function affects the simulation results very slightly (see tables IV and V).

The project will continue with measuring, modeling and simulation of the Change Management processes of another telecommunication service provider to approve and generalize our results.

All of these results will be further published and presented.

ACKNOWLEDGMENT

This work has been supported by the grant FRVŠ G1 2030/2011 given by the Czech Ministry of Education, Youth and Sports.

REFERENCES

[1] R. Luecke, Managing Change and Transition. 1st ed., Harvard Business Press. 2003. 138 p. ISBN: 1578518741.

[2] J. Hiatt, T. Creasey, Change Management. 1st ed. Prosci Research. 2003. 148 p. ISBN: 1930885180.

[3] B. Melamed, Simulation Modeling and Analysis with ARENA. 1st ed. Academic Press. 2007. 456 p. ISBN: 0123705231.

[4] J. Rayner, O. Thas, D. Best, Smooth Tests of Goodness of Fit: Using R. 2nd en. Wiley. 2009. 304 p. ISBN: 0470824425.

[5] C. Huber-Carol, N. Balakrishnan, M. Nikulin, M. Mesbah, Goodness-of-Fit Tests and Model Validity: Statistics for Industry and Technology. 1st ed. Birkhäuser Boston. 2002. 507 p. ISBN: 0817642099.

[6] R. Ash, Basic Probability Theory. 1th ed. Dover Publications. 2004. 352 p. ISBN 0486466280.

[7] J. Devore, Probability and Statistics for Engineering and the Sciences. 7st ed. Duxbury Press. 2008. 768 p. ISBN 0495557447.

Traffic Analysis in Contact Centers

Erik Chromy, Jan Diezka, Matus Kovacik, Matej Kavacky
Institute of Telecommunications, Faculty of Electrical Engineering and Information Technology,
Slovak University of Technology, Ilkovičova 3, 812 19 Bratislava, Slovak Republic
chromy@ut.fei.stuba.sk, jan.diezka@gmail.com, kovaci@ut.fei.stuba.sk, kavacky@ut.fei.stuba.sk

Abstract— **The paper deals with the contact centers with the emphasis on traffic analysis, monitoring of significant parameters and way of modeling of this traffic during contact center design. The main goal of this work is to show the utilization of Erlang traffic theory for simulation and calculation of essential traffic parameters of contact centers. The term traffic means especially the processing of phone call in this case in contact center.**

Keywords-Contact Center; Erlang B formual; Erlang C formula; QoS

I. INTRODUCTION

Today the margin between success and failure in business is very tiny. Continuous contact with customers plays the key role of success. The building of trouble free contact center helps in keeping of contacts with existing and moreover potential customers. Contact center (CC) is complex communication system serving as communication medium between corporate and their customers. It is a software and hardware superstructure of PBX, which is main gateway for voice communication for organization. CC combines technical components with human factors, e.g. agents of CC, managers, supervisors, administrators and other employees. The complexity of CC technology allows various ways of its utilization.

CC can be used in the fields of: customer support, telemarketing, banking, insurance, information service, advertising, reservation systems and many others [2], [3], [8], [9].

II. OPERATION OF CONTACT CENTER

Corporate implementation of CC brings better approaching to customers and therefore it has a great merit for both sides. The customer has simplified access to product and service information and organization can monitor the customers' requirements and respond accordingly to improve its services. The operation of contact center brings various expenses from economic point of view. These expenses can be divided into three main groups in general: telecommunication charges, expenses for human resources and costs of technical solution [2].

The significant part of these expenses produces costs for human resources (more than 50 %) – wages, training costs, bonuses and others. These costs are periodically repeated. Hence, it is important to have the suitable number of agents in right time together with the lowest costs. The aim is to serve the largest number of customers with the lowest number of agents [2], [4].

For CC managers the effort for reducing of costs, increasing of quality and improving efficiency means the searching of answers to these following questions [5]:

- How many agents we need,
- What skills they need to have,
- How to organize their working time (shifts, breaks, lunches, trainings),
- How many, when and what type of calls we expect,
- How we can redirect call in order to the most effective utilization of agents,
- How we can predict the contact center traffic?

Answers to these questions can be found by various tools, models and simulations.

III. MODELING OF CONTACT CENTER

In general we have three contradictory and limiting traffic parameters – QoS [1], [7], satisfaction of employees and contact center costs. It is necessary to employ larger number of agents in order to improve the QoS, therefore costs are increasing. If we want to reduce the costs, the wages of agents must be decreased, if we need the same good QoS. But the wages reduction cause displeasure of employees and also quality of their work will fall. Proper dimensioning of contact center means to find the system in which all these parameters will be balanced [5].

The operation of contact center is based on the following principle: customers are contacting contact center with various requests. The first contact with customer is done by IVR (interactive voice response) which is responsible for customer identification and it is capable of processing of some customer requests. If the request processing needs the help of agent, IVR will redirect the call to ACD (automatic call distribution) module which will connect the call to the most appropriate agents through special routing algorithms. If the number of callers is higher than number of agents, waiting queues will rise. ACD also manages these waiting queues. Finally, the contact center agent will handle the customer request.

The customers call randomly to contact center. Some of them will wait on the line through whole process (contact with

IVR, redirecting by ACD and waiting in the queue), others will hang up after contact with IVR or during waiting in the queue.

If we want to have a fully functional model of contact center, it is important to precisely simulate its traffic. It is necessary to define basic parameters, that will ease monitoring of contact center traffic:

- Percentage of successfully answered calls,
- Percentage of interrupted calls – calls aborted by customers before they were served by agent,
- Average answering speed,
- Hour of maximum traffic load,
- Average speed of call processing,
- GoS – Grade of Service – means expected level of requested services; good GoS can be for example, if 80% of incoming calls are served till 20-30 seconds (rest of the calls waits longer).

Incoming calls, from statistical point of view, are discrete random variables, their probability distribution can be well described by *Poisson probability distribution/probability density function* of discrete random variable with average value λ [6].

For traffic measuring in telecommunications we use *Erlang* unit. Hence, for traffic modeling in contact centers we can use Erlang models.

Expression "traffic" in this case means a system, where sources request specific services and servers offer these services. In contact centers, traffic sources are customers and servers are agents assigned to serve calling customers [6]. One phone terminal is served by one agent, in general.

IV. ERLANG B MODEL

One of the simplest contact center models is Erlang B model [6].

Incoming calls are modeled by Poisson distribution with average incoming value λ. All contact center links are busy with probability P_B (probability of call blocking), this means that incoming calls are considered as lost. Probability of successful call into contact center and consecutive interconnection to agent by ACD is 1-P_B. Erlang B model does not consider any waiting queues for call interconnection with agent (Figure 1).

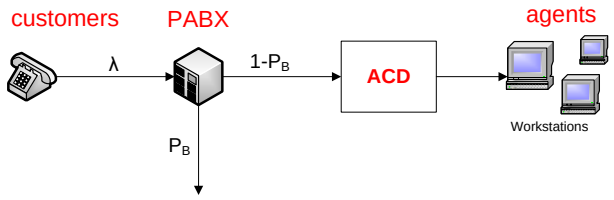


Figure 1. Erlang B model

For contact center parameter calculations by Erlang B model B we will use Erlang B formula (1):

$$P_B = \frac{\frac{A^N}{N!}}{\sum_{i=0}^{N} \frac{A^i}{i!}} \quad (1)$$

Erlang B formula uses 3 basic parameters:

- A – the traffic load in Erlang,
- N – number of agents /lines / trunks (requested simultaneous connections),
- P_B – probability of call blocking.

Thanks to Erlang B formula we can calculate unknown third parameter if we know the remaining two. For successful contact center it is very important to define exact number of agents (N), needed for handling all of the contact center traffic (A).

A. *Calculating the right number of agents*

Let's suppose that the model for calculation of number of agents is following: modeled contact center works with traffic load of A = 5 Erl. This load equals average of 60 incoming calls in one hour, average serving time is 5 minutes (cca 0,084 hour) and contact center works 8 hours a day.

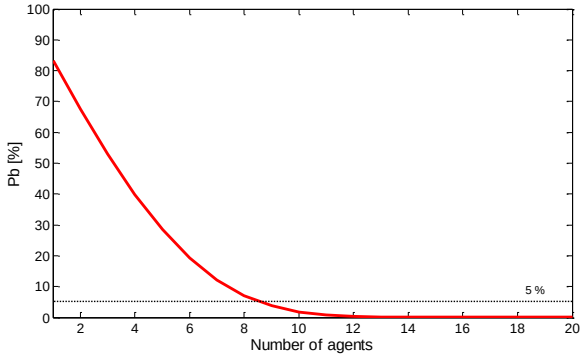


Figure 2. Relation between P_B and number of agents if traffic load A = 5 Erl.

From Figure 2 we can see that with increasing number of agents, probability of call rejection P_B decreases sharply till particular level. After reaching 5 % this sharply decreases stops. In this case, it is after reaching 8 agents in contact center. After reaching 12 agents, probability of rejection P_B is almost zero. Hence, the right number of agents lies between these two values.

For deciding the exact number of agents, it is necessary to choose maximum possible probability of call rejection P_B, Contact center should not have higher call rejection probability. Then the right number of agents can be determined from following Table 1:

TABLE I. RESULTS FROM ERLANG B MODEL

A [Erl]	N	P_B [%]
5	5	28,49
5	6	19,18
5	7	12,05
5	8	7,00
5	9	3,75
5	10	1,84
5	11	0,83
5	12	0,34

In this case, maximum value of call rejection probability is P_B = 2,5 %. This value is chosen regarding to requested parameters of the contact center. It depends on individual contact center manager.

B. *Maximum load calculation with defined number of agents*

Erlang B model is also usable for maximum contact center load calculation, without changing the number of agents.

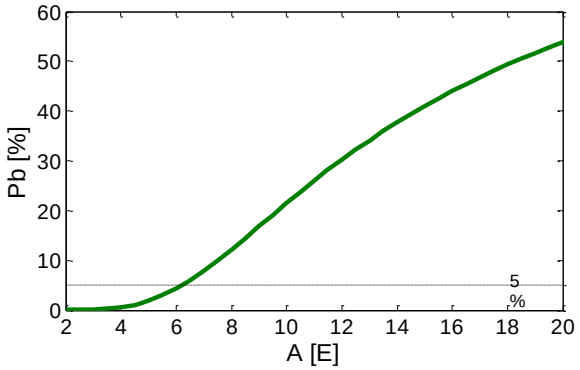


Figure 3. Relation between P_B and traffic load A with number of agents N = 10

From Figure 3 we can see that with increasing traffic load and constant number of agents, the probability of call rejection P_B sharply rises. 5 % border P_B reaches at load of 6 Erl, at 10 Erl probability of call rejection is 20 % and at 18 Erl it is 50 %, that means every second call (in average) will be rejected.

C. *Defining number of agents in relation to traffic load at constant P_B*

With increasing traffic load and constant number of agents, contact center relatively fast exceeds maximum probability of call rejection P_B. This value is individually chosen, in our analysis the maximum value of P_B is 2,5 %.

Manager of contact center really could find a chart of relation between number of needed agents N and rising traffic load A very interesting. Maximum P_B = 2,5 % , remains the same. This relation is shown in Figure 4.

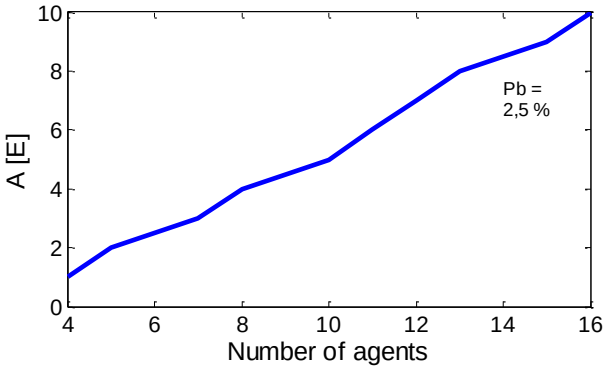


Figure 4. Relation between traffic load A and number of agents N at constant P_B=2.5%

We can see that number of agents needed to ensure such P_B is almost linear. Also Figure 4 shows that at load A = 5 Erl, for reaching maximum probability of rejection P_B = 2,5 % we need N = 10 agents, as calculated before.

D. *Result analysis of Erlang B model*

From calculated values and charts, Erlang B model can be used for calculations of following parameters of contact center:

- Number of agents needed for serving defined traffic load at certain P_B,
- Maximum possible traffic load of contact center at constant number of agents,
- Probability of call rejection at defined load and number of agents.

V. ERLANG C MODEL

Difference from B model is that in Erlang C model rejected calls are not considered as lost, but are delayed until trunk is freed and these calls are served [5], [6].

Erlang C model is considering the waiting queue with infinite waiting time. No call is lost, calling customer stays on the line till he's served (Figure 5).

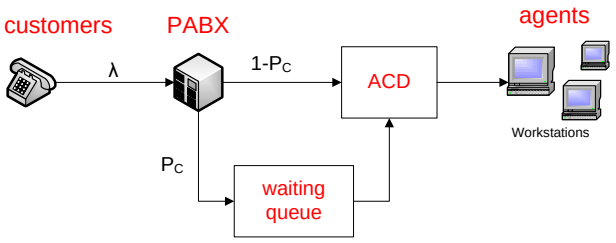


Figure 5. Erlang C model

For calculations of contact center parameters using Erlang C model, it is better to use extended Erlang C formula (2), which contains probability of call rejection defined by Erlang's B formula (1):

$$P_C = \frac{N * P_B}{N - A + A * P_B} \quad (2)$$

Formula (2) contains 4 basic parameters:

- A – traffic load of contact center in Erlangs,
- N – number of agents,
- P_B – probability of call rejection,
- P_C – probability of call waiting in waiting queue.

After defining three of these parameters, we can calculate the fourth from (2). More important parameters of contact center can be calculated if we add fifth parameter *GoS* and after combining with formula (2):

$$GoS = 1 - P_C * e^{-\frac{(N-A)t}{h}} \quad (3)$$

Formula (3) includes following important parameters:

- GoS – probability of picking up the call by agent till time *t*,
- t – maximum waiting time for call serving, fullfiling GoS condition,
- h – average call processing time by agent.

Value *t* is chosen individually by contact center manager's requirements for GoS.

A. *Calculation of required number of agents*

For exact calculation of number of agents, we need to know expected contact center traffic.

Let A = 5 Erl, which equals 60 incoming calls in one hour and average call processing time of 5 minutes.

Probability of call rejection P_B should be maximally 5 % and probability of waiting in queue for call serving also 5 %, so that only 3 calls wait in the queue from total number of 60 incoming calls.

Figure 6 we can see that with increasing number of agents, probabilities P_B and P_C decrease. Probability P_B reaches the chosen value at 9 agents, probability P_C at 10 agents. Exact numbers are shown in Table 2.

From Table 2 it is evident that at 10 agents GoS is almost 98 %, meaning that only 2 % of incoming calls are waiting in queue for longer time than 30 seconds. By increasing number of agents to 11, GoS rises to 99%, that is only 1 % raise, but expenses for that one agent would raise much more than 1 %. Hence, from these calculations we can say that addition of eleventh agent is not necessary.

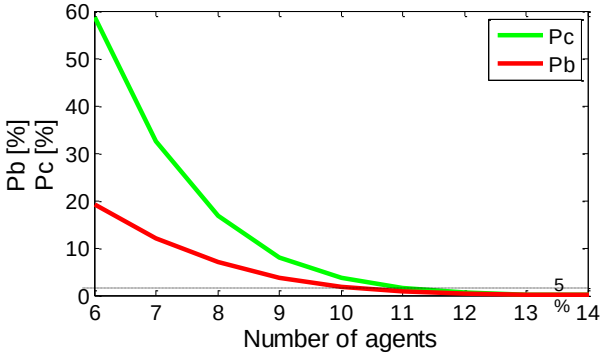


Figure 6. Relation of P_B and P_C from the number of agents at traffic load A = 5 Erl.

TABLE II. RESULTS FROM ERLANG C MODEL

A	N	P_B [%]	P_C [%]	GoS [%]
5	6	19,18	58,75	46,83
5	7	12,05	32,41	73,46
5	8	7,00	16,72	87,60
5	9	3,75	8,05	94,60
5	10	1,84	3,61	97,81
5	11	0,83	1,50	99,17
5	12	0,34	0,58	99,70
5	13	0,13	0,21	99,90

B. *Relation between GoS and number of agents at constant traffic load*

Relation between GoS (t = 30s) and number of agents at constant traffic load is shown at Figure 7.

With increasing number of agents, GoS raises sharply, this increase stops at number of 10 agents, where GoS reaches a value of almost 98 %. Further increasing of number of agents inflicts GoS value only in minimal aspect.

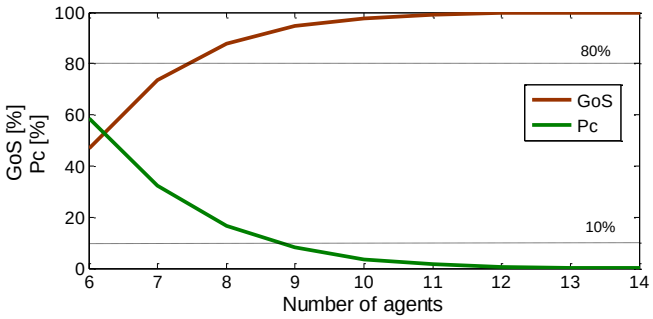


Figure 7. Relation of *GoS* and P_C to number of agents at load A = 5 Erl

80 % value reaches GoS already at 8 agents; in this case the probability of call redirection to waiting queue is more than 10 %.

At 9 agents both values are good for contact center performance. From Figure 7 is evident that GoS relation to number of agents is opposite to relation between probability of waiting in queue P_C and number of agents.

C. *Contact center stability*

Contact center stability means performance level of contact center at defined traffic load. This level is indicated by different parameters. Most important parameter is GoS value at chosen *t* parameter, other parameters are probability of call redirection to waiting queue P_C, number of calls waiting in the queue to be served and probability of call rejected due to link blocking P_B.

Contact center manager's work is to choose values as limits for these parameters.

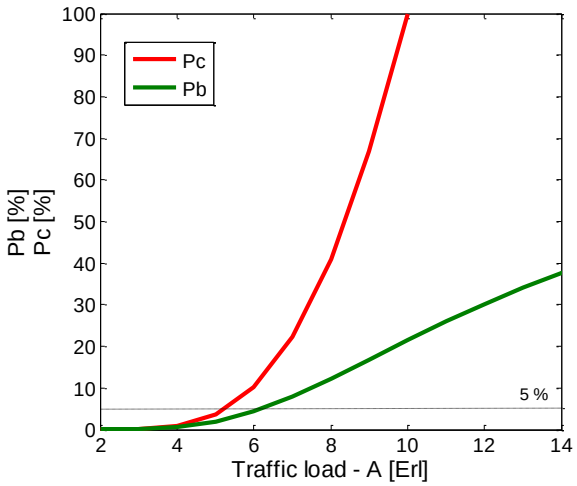


Figure 8. Relation between P_B and P_C and traffic load A at number of agents $N = 10$

For traffic load of 5 Erl, calculated number of agents needed to handle such load is 10. At this value contact center reaches GoS of 97,81 %, maximum call waiting time in queue fulfilling this GoS is 30 seconds and number of rejected calls is 1,84 % of incoming calls (will be blocked).

Form Figure 8 we can define maximum traffic load at which the contact center will perform well. It's obvious that with increasing traffic load the probability of call redirection to waiting queue P_C raises sharply. Already at load of 5 Erl, P_C is more than 5 %, at load of 10 Erl, the contact center is full – probability of call redirection to waiting queue is 100 %, and every fifth call is blocked.

D. *GoS and increasing load influences on average call serving time*

Resulting from previous analysis done on an model example, for good performance of contact center at traffic load of A = 5 Erl we need N = 10 agents, in case that the average call serving time by agent is h = 5 minutes (300 seconds). Values *A* and *h* were chosen and number of agents was then calculated from Erlang C model. With this parameters, contact center works with GoS value of 97.81 % , fulfilling condition of t = 30 s.

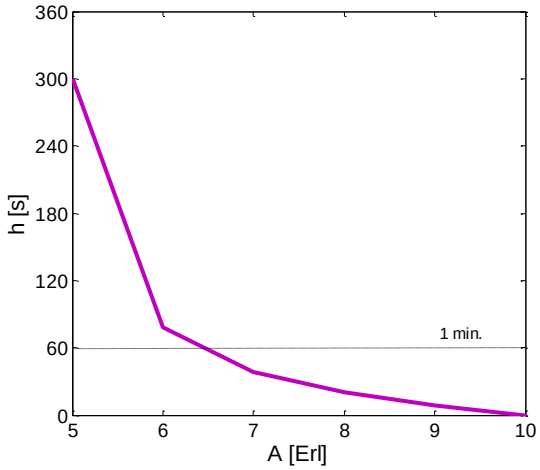


Figure 9. Relation between average call serving time and traffic load *A* at the same value of *GoS*

Contact center's traffic load increases from starting value of A = 5 Erl (60 incoming calls per hour), average call serving time is h = 5 minutes; at this load.

From Figure 9 it's evident that we need to lower the value of average call time with increasing number of incoming calls. For keeping the same GoS at increasing traffic load, agents have to work much faster. Only 1 Erlang increase in traffic load causes almost untenability of GoS at the same number of agents.

E. *Result analysis of Erlang C model*

From calculated values and charts, Erlang C model (after extension by GoS parameter and modification of Erlang C formula by Erlang B formula) can be used for calculations of following parameters of contact center:

- Number of agents needed to handle traffic load *A* at defined maximum values for P_C a P_B,
- GoS at defined number of agents, defined traffic load and average call serving time,
- Maximum possible traffic load at defined number of agents (contact center stability),
- Maximum average call serving time in relation to GoS value,
- Probability of call rejection (P_B), at defined load and number of agents,
- Probability of call redirection to waiting queue (P_C), at defined load and number of agents.

VI. CONCLUSIONS

It was shown that Erlang B model prerequisites for contact center are not sufficient (no existing waiting queue), Erlang B model is not optimal for calculations of number of needed agents. But this model can be used for parameter calculations through Erlang C model, where it can calculate probability of

link blocking or the number of needed phone lines in the contact center.

Erlang C model analysis shows that it can be used for calculations of all necessary and important traffic parameters of a contact center. Erlang C model's main deficiency is the calculation with infinite waiting queue and customer's infinite patience. Other deficiency is the lack of IVR consideration in contact center.

Advantage of Erlang C model is in simplicity and wide scale of calculations for almost all important traffic parameter of a contact center.

ACKNOWLEDGMENT

This work is a part of research activities conducted at Slovak University of Technology Bratislava, Faculty of Electrical Engineering and Information Technology, Department of Telecommunications, within the scope of the projects VEGA No. 1/0565/09 „Modelling of traffic parameters in NGN telecommunication networks and services" and ITMS 26240120029 "Support for Building of Centre of Excellence for SMART technologies, systems and services II".

REFERENCES

[1] Baroňák, I., Halás, M.: Mathematical Representation of VoIP Connection Delay. In: Radioengineering. ISSN 1210-2512. Vol. 16, No. 3 (2007), pp. 77-85.

[2] Baroňák, I., Chromý, E.: Contact center – part of modern communication infrastructure, In: Telekomunikace, no. 11, November, 2004, pp. 22-26.

[3] Anisimov, N., Kishinski, K., Miloslavski, A.: Formal Model, Language and Tools for Design Agent's Scenarios in Contact center Systems, In: 32nd Hawaii International Conference on System Sciences, 1999

[4] Shen, H., Huang, J., Lee, Ch.: Forecasting and Dynamic Updating of Uncertain Arrival Rates to a Contact center, In: IEEE International Conference on Service Operations and Logistics, and Informatics, Soli, 2007

[5] Hishinuma, Ch., Kanakubo, M., Goto, T.: An Agent Scheduling Optimization for Contact center, In: The 2nd IEEE Asia-Pacific Services Computing Conference, 2007

[6] DIAGNOSTIC STRATEGIES: Traffic Modeling and Resource Allocation in Contact centers (*www.DiagnosticStrategies.com*) *Diagnostic Strategies, 371 Country Way, Needham, MA 02492*

[7] Baroňák, I., Kvačkaj, P.: A New CAC Method Using Queuing Theory. In: Radioengineering. ISSN 1210-2512. Vol. 17, No. 4 (2008), Part. II, pp. 62-74.

[8] Baroňák, I., Ferenczyová, I.: Implementation of Contact Centre to Healthcare Sector. In: ICETA 2008: 6th International Conference on Emerging e-Learning Technologies and Applications. High Tatras, Slovakia, September 11-13, 2008. Košice: Elfa, 2008. ISBN 978-80-8086-089-9.

[9] Mišuth, T., Baroňák, I.: Performance Forecast of Contact Centre with Differently Experienced Agents. In: Elektrorevue. ISSN 1213-1539. Vol. 15, 16.11.2010 (2010), art. no. 97.

[10] Balogh, T., Luknárová, D., Medvecký, M.: Performance of Round Robin-Based Queue Scheduling Algorithms. In: CTRQ 2010: The Third International Conference on Communication Theory, Reliability, and Quality of Service. Athens, Greece, 13.-19.6.2010. Los Alamitos: IEEE Computer Society, 2010. ISBN 978-0-7695-4070-2, pp. 156-161

Controlling Asterisk with Java Application

Pavel Nevlud, Lukas Kapicak, Jaroslav Zdralek, Jan Plucar, Patrik Dubec
Department of Telecommunications
VSB-Technical University of Ostrava
Ostrava-Poruba, Czech Republic
pavel.nevlud@vsb.cz, lukas.kapicak@vsb.cz, jaroslav.zdralek@vsb.cz, jan.plucar@vsb.cz, patrik.dubec@vsb.cz

Abstract— **The paper deals with remote control of the Asterisk with java application. We developed Java application for remote control of the Asterisk through the Asterisk Manager Interface. We used the java libraries Asterisk-Java and we wrote the application for sending any CLI command with the java-based program. We developed java-based interface for controlling Asterisk through SSH protocol.**

Keywords: Asterisk, Asterisk-Java, Asterisk Manager Interface, NetBeans, CLI, SSH

I. INTRODUCTION

In the INDECT Project [1] we participate in the work package 6, that is developing the Web Portal. We prepare part of portal that will be used for mobile end users devices. We cooperated on developing Interactive Video Application System IVAS that is based on the open-source telephony exchange called Asterisk. We developed Java application for remote control and managing the Asterisk. As integrated developing environment we had chosen NetBeans, because it is free, open source software with a lot of features [5].

II. REMOTE CONTROL OF ASTERISK

Asterisk is software that may turns personal computer into a voice communications server. Asterisk is powerful and popular telephone development tool-kit. It may be used by small business, large businesses, call centers, carriers, governments worldwide etc. We used Asterisk as IVVR (Interactive Voice and Video Response) system.

One of the biggest advantages of Asterisk is the ability to external remote control from shell or through external applications. We tested some possibilities for remote controlling of Asterisk. In this part we describe our experience with various possibilities.

A. Remote control through secure shell

The simplest way to control Asterisk from an external shell or application is to issue the command with option -rx.

```
asterisk -rx "dialplan show 100@default"
```

Where Asterisk process CLI command `dialplan`. This command shows extension 100 in the default context.

Next example shows how to create call between SIP user and extension number 100. Connecting to the extensions we are able to perform such new functions. Between these functions belongs establishing audio and video calls, video recording, video playback or connecting users to remote IP camera.

```
asterisk -rx "originate SIP/user extension 100"
```

B. Remote control through "Call Files"

The second possible way is to use "Call Files" which are like a shell script for Asterisk. User or application writes a call file into Asterisk specific directory, and then Asterisk immediately will process it.

For an example we prepared file call_user.call and moved this file into specific Asterisk directory. The information about the date and time is very important. Asterisk will processes the file when time in file is same as the time of Asterisk.

```
mv call_user.call /var/spool/asterisk/outgoing
```

The file call_user.call may content following information:

```
Channel: SIP/user
Priority: 1
Context: default
Extension: 100
```

C. Remote control through the Asterisk Manager Interface

The next access for remote control may be realized through activated Asterisk Manager Interface AMI.

In a file `/etc/asterisk/manager.conf` we have to create new entry with "AMI user" for activating AMI interface. We created admin user with password secret and we permit remote control only from one IP address. The address was our local address so we can be connected to Asterisk only from our station.

```
[manager]
secret = xxxxx
deny = 0.0.0.0/0.0.0.0
permit = 127.0.0.1/255.255.255.255
read = all
write= all
```

After reloading `manager.conf` or restarting Asterisk we may establish connection with Asterisk through AMI interface on the port 5038 from the system shell using telnet. For our testing purpose we connected to Asterisk from local IP address.

```
$ telnet 127.0.0.1 5038
Trying 127.0.0.1...
```

```
Connected to localhost.
Escape character is '^]'.
Asterisk Call Manager/1.0
```

Next step of our configuration is use of action command for authentication process. After successful authentication we can use another action command for control Asterisk.

```
action: login
username: manager
secret: xxxxx
```

For insert our command we have to use the Enter button twice to send this action to the Asterisk. At the end we must use to action logoff for logout from the Asterisk. This action command is implemented in the Asterisk-Java package.

D. Remote control through SSH

The Asterisk Manager Interface enables applications to connect to Asterisk server instance and issue various commands. Manager provides several actions, however the important one is "Command" action. This action can be used to define any CLI command. Here is an example from our application:

```
   private static final String cmdConnectUsers =  "originate
local/&user1&groupDef1" + "extension" + "&user2&groupDef2";

   private String ConnectUsers(String caller, String
callerContext, String called, String calledContext)
   {
       String command = cmdConnectUsers.replace("&user1",
caller);
       command = command.replace("&user2", called);
       command = command.replace("&groupDef1",
callerContext);
       command = command.replace("&groupDef2",
calledContext);

       CommandAction cmdAction = new CommandAction(command);
           //Code handling connection
          return result;
   }
```

We have decided to use web services to propagate methods for public use. Anyone can create client application and use our web service to control asterisk server without worrying about functionality behind this web service methods. We can also make changes to the code without forcing client applications to be changed accordingly.
Our solution uses multilayered architecture in order to achieve easier code manipulation and changes. We have divided project into three layers:

- Connection Layer,
- Administrative Layer
- Presentation Layer.

Connection Layer handles connection to server; issue commands and parse the results. Administration Layer contains definitions of Status codes, Exceptions and Entities. Presentation Layer returns parsed data.

III. ASTERISK-JAVA

The Asterisk-Java [3] package consists of a set of Java classes [6] [7] that allow build Java applications that interact with Asterisk. Asterisk-Java supports many features of Asterisk interfaces like Fast Asterisk Getaway Interface, Asterisk Manager Interface.

Asterisk-Java is provided under the terms of the Apache License, Version 2.0. We used release Asterisk-Java 1.0.0.M3, that is compatible with Asterisk 1.6.

A. The Manager API

The Manger API is interface for remote interaction with Asterisk. The Manager API allows query and change its state at any time. The manager API is made up of three concepts: Action, Responses and Events.

Actions commands can be sent to the Asterisk and instruct it to do some specific function. For example application can send action command to the Asterisk requesting it to dial a number and direct the dialed party to one extension, like interactive voice response. In reply to an action command Asterisk sends a Response that contains the results of the operation performed.

Events are sent by Asterisk without a direct relation to the Actions which your application is sending. Events inform you about the relevant changes in Asterisk's state. For example Events are used to inform your application about incoming calls or registered peers.

The connection to the Asterisk server via Manager API occurs over TCP/IP usually on the default port 5038. To enable the Manager API in Asterisk you must edit `manager.conf` configuration file and restart or reload Asterisk services.

Asterisk-Java automatically reconnects if it loses the connection to Asterisk.

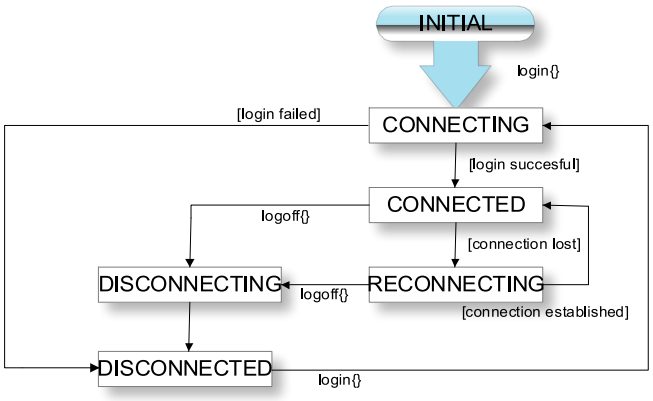


Fig. 1 Flowchart for automatically reconnection

B. Debugging Manager API

For debugging purpose is very useful and important tool **ngrep**. We can start ngrep on the Asterisk server or on the client. A successful login results are in output:

```
ngrep -d lo -s 1500 port 5038 -T

interface: lo (127.0.0.0/255.0.0.0)
filter: (ip or ip6) and ( port 5038 )
####
T +13.188292 127.0.0.1:5038 -> 127.0.0.1:53695 [AP]
 Asterisk Call Manager/1.0..
##
T +0.038662 127.0.0.1:53695 -> 127.0.0.1:5038 [AP]
```

```
 action:  Challenge..actionid:  2989062_0#..authtype:
MD5....
##
T +0.000048 127.0.0.1:5038 -> 127.0.0.1:53695 [AP]
 Response:  Success..ActionID:  2989062_0#..Challenge:
144063931....
##
T +0.009741 127.0.0.1:53695 -> 127.0.0.1:5038 [AP]
 action:    Login..actionid:    2989062_1#..authtype:
MD5..username: manager..key:
   a5fdfffc0541d75e6d740f56874164b1....
#
T +0.000146 127.0.0.1:5038 -> 127.0.0.1:53695 [AP]
 Response: Success..
#
T +0.000010 127.0.0.1:5038 -> 127.0.0.1:53695 [AP]
 ActionID: 2989062_1#..
##
T +0.000047 127.0.0.1:5038 -> 127.0.0.1:53695 [AP]
 Message: Authentication accepted....
```

In this log we can see successful output comes from Asterisk. If appear some problems we may immediately correct it.

IV. Asterisk Java Applications

A. *IDE NetBeans*

We have chosen IDE Netbeans [8] [9] as development tool, because it is free, open source Integrated Development Environment with a lot of features. In the future we are planning to change our application into web services that will be deployed on the GlassFish server.

Next figure shows NetBeans IDE with Asterisk CLI java application. This application can send any CLI command to Asterisk and appropriate response we can see as output.

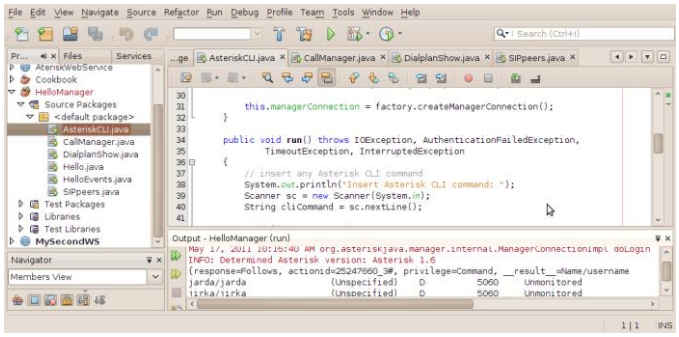


Fig. 2 NetBeans IDE with Asterisk CLI java application

B. *SIP peers Java application*

First we have to insert asterisk-java.jar into Netbeans libraries [3] and then we can create our application. Now we are able to write an application for getting some information from Asterisk with Action SipPeers.

```
import java.io.IOException;
import org.asteriskjava.manager.AuthenticationFailedException;
import org.asteriskjava.manager.ManagerConnection;
import org.asteriskjava.manager.ManagerConnectionFactory;
import org.asteriskjava.manager.ManagerEventListener;
import org.asteriskjava.manager.TimeoutException;
import org.asteriskjava.manager.action.SipPeersAction;
import org.asteriskjava.manager.event.ManagerEvent;

public class SIPpeers implements ManagerEventListener {
  private ManagerConnection managerConnection;

  public SIPpeers() throws IOException  {
      ManagerConnectionFactory factory = new
ManagerConnectionFactory("IP_Asterisk", "manager", "xxxxx");
      this.managerConnection = factory.createManagerConnection();
      }

      public void run() throws IOException,
AuthenticationFailedException,
              TimeoutException, InterruptedException  {
        managerConnection.addEventListener(this);
        managerConnection.login();
        managerConnection.sendAction(new SipPeersAction());
        Thread.sleep(1000);
        managerConnection.logoff();
       }
       public void onManagerEvent(ManagerEvent event)  {
        String sablona = "ipaddress=[null0-9'.]*,timestamp=[null0-
9.]*,
              objectname='[A-Za-z0-9]*'";
        Pattern pattern = Pattern.compile(sablona);
        String toString = event.toString();
        Matcher matcher = pattern.matcher(toString);
        if (matcher.find()) {
           System.out.println(matcher.group());
        }
        else
           System.out.println();
       }

       public static void main(String[] args) throws Exception  {
        SIPpeers sippeers;
        sippeers = new SIPpeers();
        sippeers.run();
       }
    }
```

Next figure represented how we tested our Java applications.

The Java application runs on the other computer than Asterisk. We remotely control Asterisk through Java application with invoke Action SipPeers. In our testing we used Asterisk 1.6.2.5 version.

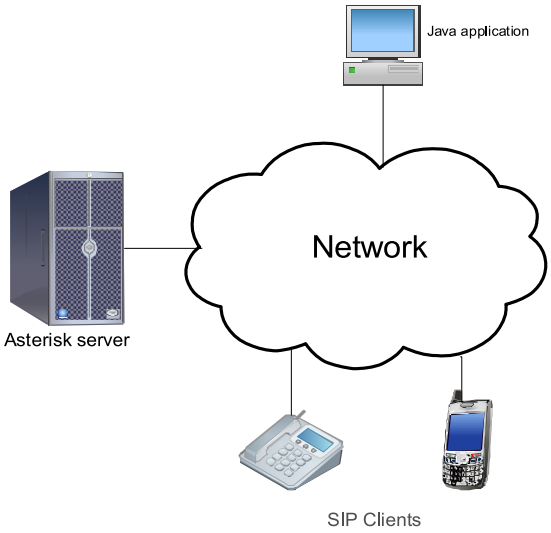


Fig. 3 Remote control of Asterisk through Java application

When we run this SIPpeers aplication we get response from Asterisk server that looks like next output. Our program filter output information with using regular expressions. We can see that user user3 is not registered to Asterisk server.

```
ipaddress='192.168.1.100',timestamp=null,object
name='user1'
ipaddress='192.168.1.151',timestamp=null,object
name='user2'
ipaddress=null,timestamp=null,objectname='user3
'
```

C. CLI Command Java application

CLI commands managed by Java application were the same Asterisk-Java package that was inserted in NetBeans. This application use Action Command to allow using any command from Asterisk.

```
    import java.io.IOException;
    import java.util.Scanner;

    import org.asteriskjava.manager.AuthenticationFailedException;
    import org.asteriskjava.manager.ManagerConnection;
    import org.asteriskjava.manager.ManagerConnectionFactory;
    import org.asteriskjava.manager.ManagerEventListener;
    import org.asteriskjava.manager.TimeoutException;
    import org.asteriskjava.manager.action.CommandAction;
    import org.asteriskjava.manager.event.ManagerEvent;
    import org.asteriskjava.manager.response.ManagerResponse;

    public class AsteriskCLI implements ManagerEventListener  {
        private ManagerConnection managerConnection;

        public AsteriskCLI() throws IOException  {
         ManagerConnectionFactory factory = new
ManagerConnectionFactory(
                    "IP_Asterisk", "manager", "xxxxx");

          this.managerConnection = factory.createManagerConnection();
        }

        public void run() throws IOException,
AuthenticationFailedException,
                TimeoutException, InterruptedException  {
          // insert any Asterisk CLI command
          System.out.println("Insert Asterisk CLI command: ");
          Scanner sc = new Scanner(System.in);
          String cliCommand = sc.nextLine();

          CommandAction commandAction;
          ManagerResponse commandResponse;
          commandAction = new CommandAction();
          commandAction.setCommand(cliCommand);

          managerConnection.addEventListener(this);
          managerConnection.login();
          managerConnection.sendAction(commandAction);
          Thread.sleep(1000);
         commandResponse=managerConnection.sendAction(commandAction);
          System.out.println(commandResponse.getAttributes());
          managerConnection.logoff();
        }

        public void onManagerEvent(ManagerEvent event)  {
          System.out.println(event);
        }

        public static void main(String[] args) throws Exception  {
          AsteriskCLI asteriskCLI;
          asteriskCLI = new AsteriskCLI();
          asteriskCLI.run();
        }
    }
```

When we run this java application AsteriskCLI, we can insert any Asterisk CLI command and we get response from Asterisk server. We may for example create call between sip clients with command:

`originate sip/user extension 100.`

Asterisk executes this command and creates connection between **user** and number **100**. And we can see events output from Asterisk in the Netbeans IDE.

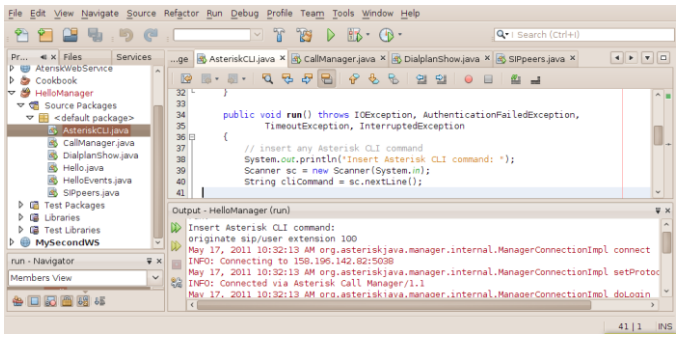


Fig. 4 Output from Netbeans IDE

V. Conclusion

This paper presented our experiences with remote control of Asterisk via Manager API. We used java library asterisk-java and we created own application for get various state information from Asterisk and send control commands to Asterisk. We have chosen JAVA language because it is easy to connect and deploy our solution as a web services. In the near future we plan extend our Java applications into web services and offer to other application control Asterisk through SOAP protocol. Web services clients will be able to control Asterisk with specific commands.

Acknowledgment

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 218086.

References

[1] INDECT Project http:/www.indect-project.eu

[2] Jim van Meggelen, Jared Smith, Leif Madsen Asterisk: The Future of Telephony. O'Reilly Media, 2 edition February 2009,

[3] Stefan Reuter Asterisk-Java, The free Java library for Asterisk PBX integration [online]. [cit.2011-03-18] http://asterisk-java.org/

[4] Ian F. Darwin Java Cookbook, Second Edition, O'Reilly 2004

[5] Netbeans Integrated Development Environment www.netbeans.org

[6] Y. Daniel Liang, "Introduction to Java Programming", Comprehensive (8th Edition), Prentice Hall, 2010, pp. 485-491.

[7] Stuart Reges, Marty Stepp, "Building Java Programs," Addison Wesley, 2. Edition, 2010, pp. 871 – 888.

[8] Jurgen Petri, "NetBeans Platform 6.9 Developer´s Guide," Packt Publishing, 2010, pp. 145-169.

[9] Tim Boudreau, Jesse Glick, Simeone Greene, Jack Woehr, Vaughn Spurlin, "Netbeans: The Definitive Guide," O´Reilly Media, 1. Edition, 2002, pp. 321-352.

[10] Heiko Boeck, "The Definitive Guide to NetBeans Platform," Apress, 1. Edition, 2009, pp. 101-108.

Management of Routing Using Artificial Intelligence

Vladislav Skorpil and Roman Precechtel
Dept. of Telecommunications, BUT Brno
Faculty of Electrical Engineering and Communication
Brno, Czech republic
skorpil@feec.vutbr.cz
precechtel@seznam.cz

Abstract - The aim of this paper is to create an active network element with the logic of routing the switching operation by using artificial neural network, and to analyze and test the possibilities of using such networks for the problem solution. The tool used for testing and analysis is MATLAB-SIMULINK with specialized extension Neural Network Toolbox for neural networks. The subject of testing is forward neural networks. The requirement is to create logical routing with artificial intelligence, at least as efficient as that implemented currently via digital circuits. The most difficult part of the whole research is network training for the required accuracy and the bulk of the work is devoted to this issue.

Keywords - Artificial Neural Network; Network Element; MATLAB

I. INTRODUCTION

The work is based on [1]. An important part of communication and information networks is active network elements that allow you to route data from one place to another. Some services have a higher priority than others, in particular services in real time. The best-known example is the transfer of voice over the Internet (VoIP). These and other types of traffic make ever heavier demands on active network elements, which must provide correct routing, sufficient bandwidth, and also ensure the required parameters for the connection of the quality of services (QoS). This is one of the main reasons for the development of telecommunications and permanent search for new ways of increasing their throughput and efficiency of [9].

Another important point of development is, for example, the increasing routing tables according to which the routers determine the path for each link [13]. The essence of active network elements is to provide a link for each device in the network and allow their mutual communication. Network elements transmit data units from the sender to the recipient on the basis of processing the results in the control unit of the network element [10]. A simplification of the process of processing the packets would be the omission of the requirements of the priorities for the management of individual requirements. It is requirement for QoS that makes great demands on the logical management of active control elements of the network. There are several options how to solve the problem of optimizing the management of the active element. This is a very complex and challenging task, with several conflicting requirements such as speed of processing the request, response to changes in network topology, memory size, price, etc [11].

Given the complexity of the problem, an appropriate solution method is processing via multiple parallel processes. This method is, inter alia, possible for and also characteristic of neural networks, both organic and artificial. The latter, the case of artificial neural networks, is the subject of this work, [12]. This differentiates it from the classical approach, which is based on the sequential processing of incoming data [12].

II. REALIZATION OF AN ACTIVE NETWORK ELEMENT

For the implementation of the management of the active element, programmable logical fields are reckoned with, which are formed by a field of programmable cells [2]. When they are connected appropriately, they provide the required logical function while the subsequent hardware implementation allows creating a specific active element capable of controlling the operation of the network [3]. The logical schema of such a connection, according to which a model would be created for processing in the form of a chip, is the subject of this work and suitable simulations leading to an optimal solution.

III. MANAGEMENT OF ROUTING USING ARTIFICIAL INTELLIGENCE

The aim of this project is to replace the controlling logical unit inside an active network element by a neural network [4]. Fig. 1 shows the basic schema of individual internal blocks. Input ports transmit incoming data to a central memory, which can have different forms of data processing or various types of queues [5]. The type of queue is one of the essential parameters of the proposal of logical routing management. Here, the basic model is FIFO (First In First Out), namely for each input port a separate queue [6].

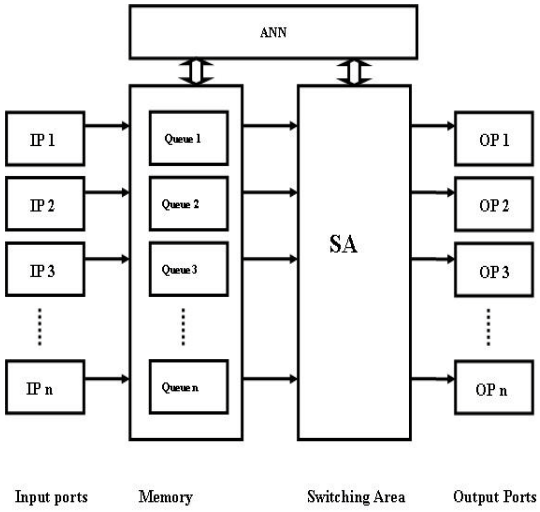


Fig. 1: Router block diagram.

IP-Input Ports, OP-Output Ports, SA-Sampling Area, ANN — Artificial Neural Network, Pr-priorities

IV. ILLUSTRATIVE EXAMPLE OF THE PRINCIPLE OF ROUTING WITH PRIORITIES

The router receives packets at entry ports 1, 2 and 3. The packets are marked 112, 112, 112 in the first queue, 214, 211 in the second queue, and 311, 322, 322 in the third queue. The first number is the input port, the second is the number of the output port, and the third number gives the priority. The highest priority is 1. Packets from the first port thus have an address that corresponds to output port 1 and priority 2, the second highest. At the same time there is in the queue of the second input port also a packet with the destination address port 1, but with priority 1, i.e. the highest priority. On the third input port there is, at the same time as the previous two packets, a packet of priority 2 addressed to output port 2. Other ports at the same time detect no traffic.

Fig. 2 gives the central memory with packets. There is a collision on outgoing port number 1, when at the same time two different packets with different priorities are bound for this port. It is also possible that two or more packets with the same priority are on one output port at the same time. In this case there are again a number of options how the control unit can respond [5]. One of the variants is the random selection and subsequent switch, possibly also a decision based on another parameter such as the similarity of packets from the previous switch, etc. A packet collision on a single default port is shown in Fig. 3.

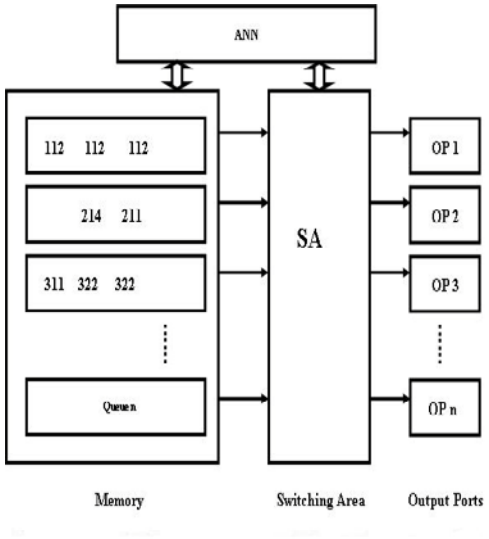


Fig. 2: Central memory with packets.

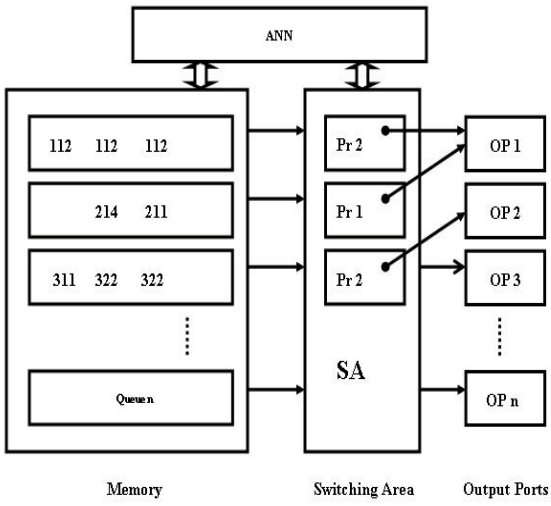


Fig. 3: Collision on outgoing port with different priority.

It is desirable for the control unit in the form of ANN to be able to evaluate this situation as follows: the packet from input port 2 with priority 1 and the packet from input port 3 with priority 2 will be switched and their queue will be shifted by one packet. This is indicated by arrows in Fig. 4. The packet from queue number 1, which remained non-switched, will have, thanks to lower significance priority, the priority number reduced by one degree, as shown in Fig. 4 in the first queue, and the evaluation of the whole process is repeated. In the next step the packets from queue 1 and 3 will be switched.

The packet from queue 2 has this time priority 4 and must therefore give priority to a packet with higher routing priority (in this case, 1is the highest priority for routing and takes priority over the others). At the same time, priority 4 is again reduced to 3 in the packet from queue 2, which had to give priority to the packet from queue 1. In the next instant of time, unless another packet comes, the two remaining packets will be routed simultaneously, since

there is no collision on the output port and priority will in this case play no role.

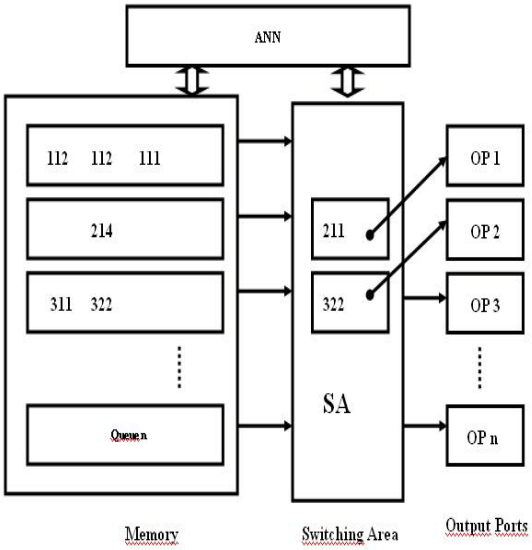


Fig. 4: Switch of packets with a higher priority and reduction of priority of non-switched packet

V. SOURCE & TARGET TRAINING DATA

The most important issue for an overall routing management design of the active nodal control element is the proposal of logical control unit representing ANN and its so-called training [6]. Training is a process in which we place on the network input the so-called training input and simultaneously output data, or the required data, i.e. the desired response to the entry of ANN. By the so-called training process ANN is trying to reach a consensus between the specified output data and the result of the calculation after the passage of training data through the network [7], in other words, it wants to learn the transfer function of the system. For a successful training it is therefore necessary to choose not only a correct ANN and its setting, but also appropriate training data [8] for ANN to be able to correctly set the network parameters or the priorities. However, even so we have no guarantee that the result will be acceptable; it will move in the range of permissible deviations from the required value. Table 1 shows examples of training data. Fig. 5 + Fig. 6 show the graphical results of MMATLAB simulation.

Index	IP	OP	Pr
0	1	1	1
0	1	1	2
0	1	1	3
1	1	2	1
1	1	2	2
1	1	2	3
2	1	3	1
2	1	3	2
2	1	3	3
3	1	1	1
3	1	2	2
3	1	3	3
4	2	1	3
4	2	2	2
4	2	3	1
5	2	1	2
5	2	2	1
5	2	3	3

Table 1 An example of training data.
IP – Input Port, OP – Output Port, Pr – Priority

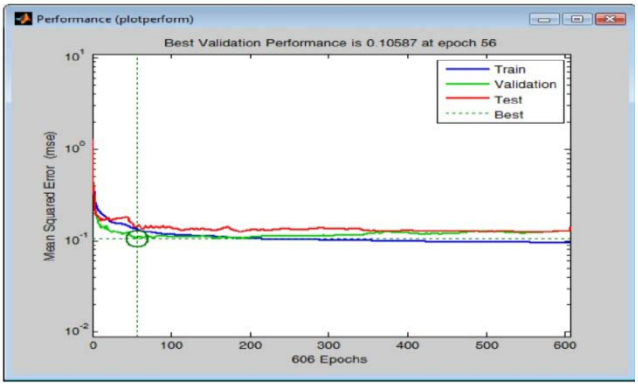


Fig. 5 Feed-forward back-propagation network

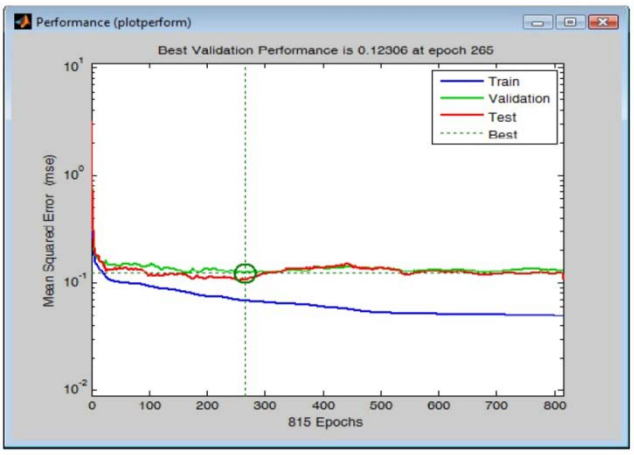


Fig. 6 Cascade-forward back-propagation network

VI. CONCLUSION

The aim of the work was to create a new architecture of the management of active nodal network element. The issue of active network elements is briefly discussed, with the emphasis on routing. The above theoretical analysis is complemented with the modelling of network elements in the MATLAB-SIMULINK program. The essence of the project is to create a model of a logical management of packet-switching and supporting priority processing. In the practical part attention was given to finding the correct network settings. The entire work is an example of processing solutions in the field of artificial neural networks. The main idea in the optimization of the setting of ANN, or its training, was to create a training set with the indexes of individual training models, with forward neural networks being selected for the training. Both of these ideas were aimed at exploring the not very standard way of solving the problem and finding a completely new method of solution, which would ultimately contribute to the final solution. A new architecture has thus been proposed for the control of network elements but the proposed procedure does not yet meet the error-rate conditions. Reducing the error-rate will be the subject of further research.

ACKNOWLEDGMENT

This research was supported by the grants:
Well the MSM 0021630513 Research Plan of Brno University of Technology "of Electronic communication systems and new generation of technology (ELKOM)",
Well. FP-S-10-16 Research communication systems and networks, as Well. EN 1.07/2.2.00/15. preparation of specialists for ICT 0139-BA Teleinformatiky,
Well. EN 1.05/03/2.1.00.0072, Centre of sensor information and communication systems (SIX)

REFERENCES

[1] PRECECHTEL, r. Optimalization Drive of the Active Network Element. BUT, Brno 2009
[2] Bishop Artificial Intelligence I, II, III, IV. Academia, Prague, 1993, 1997, 2001, 2003.
[3] DANILO p. MANDIC, Recurrent neural networks for prediction, learning algorithms, architectures and stability. John Wiley, Chichester, 2001.
[4] NORGAARD, m. Neural Networks for Modelling and Control of Dynamic Systems. Springer, London 2000.
[5] DRABEK, o.; TAUFER, I.; SEIDL, p. Artificial neural networks--Theory and Application. CHEMagazin 1 (XVI), 2006, p. 12-14. ISSN 1210-7409.
[6] DRABEK, o.; TAUFER, I.; SEIDL, p. Artificial neural networks--Theory and Application. CHEMagazin 5 (XVI), 2006, p. 6-8. ISSN 1210-7409.
[7] ŠNOREK, M.; JIŘINA, m. Neural Networks and Neural-copmputers. CTU, Prague 1996.124 s. ISBN 80-01--01455-X.
[8] TUČKOVÁ, j. Introduction to the theory and application of Artificial neural networks. CTU, Prague 2003. 103s. ISBN 80-01-02800-3.

[9] DEMUTH, h.; BEALE, m. Neural Network Toolbox for Use with MATLAB. Natick (USA): The MathWorks, Inc., 1994.
[10] L.P.S. Fernandez and A.R. Ruiz and J.de J.M. Juarez, "Urban Noise Permanent Monitoring and Pattern Recognition". *Proceedings of the European Conference of Communications – ECCOM´10.* NAUN, Tenerife, Spain, Puerto De La Cruz 2010, pp. 143 - 148 ISBN: 978-960-474-250-9
[11] Susnea and a. Filipescu and v. Minzu and g., Vasil, Virtual Pheromones and Neural Networks Based Wheeled Mobile Robot Control. *Proceedings of the 13 th WSEAS International Conference on systems.* Rodos, Greece, Iceland, WSEAS, pp. 511-516, Rodos, 2009, ISBN: 978-960-474-097-0, ISSN: 1790-2769 |
[12] Bogdamov and r. Mirsu and v. Tiponut, Matlab Model for Spiking Neural Networks, *Proceedings of the 13 th WSEAS International Conference on systems.* Rodos, Greece, Iceland, WSEAS, pp. 533-537, Rodos, 2009, ISBN: 978-960-474-097-0, ISSN: 1790-2769 |
[13] Takizava and a. Fukasawa, Novel neural Network Scheme Composed of Prediction and Studies. *Proceedings of the 13 th WSEAS International Conference on systems*. Rodos, Greece, Iceland, WSEAS, pp. 611-615, Rodos, 2009, ISBN: 978-960-474-097-0, ISSN: 1790-2769 |

Dynamic Registration System of Active Network Elements in IP Telephony Laboratory

Filip Rezac, Miroslav Voznak, Jan Rozhon, Jaroslav Zdralek
Department of Telecommunications
VSB – Technical University of Ostrava
Ostrava, Czech Republic
filip.rezac@vsb.cz, miroslav.voznak@vsb.cz,
jaroslav.zdralek@vsb.cz

Martin Krcmarik
Department of Computer Science
VSB – Technical University of Ostrava
Ostrava, Czech Republic
martin.krcmarik.st@vsb.cz

Abstract— **This article describes a system for registering network active elements in laboratory of IP telephony on VSB, Technical University of Ostrava. The system keeps track of the various elements of network infrastructure and graphically maps their location in buildings, laboratories and racks. It uses a MySQL database with access via a web application created using the programming language C # and ASP.NET 4.0 technology. Access to the system have only registered users that are assigned to a role. On the basis of this role are able to access the various site of the application. The system is also able to dynamically react on changes in network elements and notify a user that is responsible for device in the network. Realized system can streamline the handling and management of network elements in working places, where software or hardware configuration is often changed.**

Keywords - The registration system, IP telephony, ASP.NET, MySQL, C#

I. INTRODUCTION

In the laboratories of IP telephony we daily work with many network elements such as routers, switches, hubs, servers, measuring equipments and many other devices which are placed in the racks. Every day there is a displacement, adding or changing hardware and software for these devices by students, professors and laboratory workers. These manipulations have to be registered not only for reasons of clarity, but mainly for easy managing of devices.

In addition, the devices may have multiple physical ports, and each of them can be configured with different IP address. Some elements, particularly servers, may also allow the creation of virtual machines. There can be many of them and each can be managed by other person. For example, in the event of failure of the physical device system needs contact all of these responsible persons. Therefore we must have information about all the virtual machines running on that device.

Our aim was therefore to design and implement a clear, simple and flexible registration system [1] which will map the distribution of network elements in laboratories. Automatic detection of disconnected or malfunctioning element also helps with clearly monitoring. In this case the registration system periodically checks the functionality of registered elements.

Registration systems for network elements are now relatively widespread, but no freely available system provides a detailed register of the devices on such range. In addition our system provides automatic detection possibilities of connected equipments. All these facts contributes to the importance of such a system, whereas the trend of continuous growth in the number of network devices is likely to continue for very long time.

II. INPUTS, OUTPUTS, USED METHODS AND TECHNOLOGIES

Users who will access the system and perform editing of individual items must be authenticated, and have assigned the appropriate role, which will represent the rights of users. Based on a user role assigned he/she can upload, edit or delete individual records. Outputs are represented as an individual summary lists containing all available information sorted by the keys which the user sets. Detailed information about the inputs and outputs are mentioned below along with a description of used technologies and method for automatically availability detection of registered elements.

A. Users Authentication and Roles

The system will work only with authorized (logged in) users. User authentication works against the internal MySQL [2] user database, however system can also be connected with authentication servers such as LDAP [3], RADIUS [4] or Shibboleth [5]. A user who is not logged in correctly does not have access to the system. Registered users have assigned one of the following roles:

- Student,
- Employee,
- Administrator.

User with the role Student is able to view only a racks in buildings and laboratories and see the elements that their contain. He can also see the virtual machines on individual

servers and IP addresses information. He can also display a table of all devices in the system.

User with assigned role Employee have rights as a user with the role Student but also he/she is able to add, edit and delete devices, their IP addresses and virtual machines from the system (Figure 1). He/she can also add a new user into the system.

User with the role Administrator will have full control over the system. In addition, Administrator is able to manage the accounts of other users - editing and deleting. Since some users in the system have not the possibility to create their own accounts, because each account can only be created by users with particular roles, it is important that any user, regardless of the assigned role in the system, is able to change their password to log into the system.

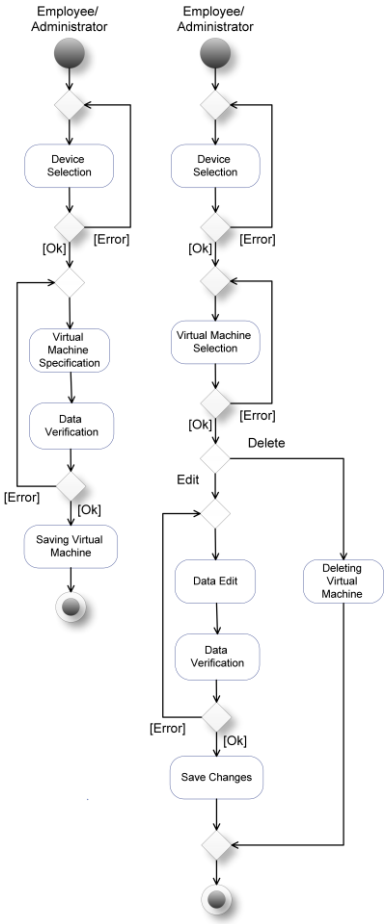


Figure 1. Activity Diagram of Creating and Editing Virtual Machine

B. *Inputs and Outputs*

For each device system keeps following information: identification number of the device, device type (such as router, server, switch, etc.). Furthermore, IP address, number of slots, which device occupies in the rack, name of the device vendor, service provided by this element, a responsible person - or a contact name, device name, adding date and time, in which building, lab, rack and position is device located, notes about the element and that device will be automatically monitored for availability.

For virtual machines system keeps: unique identification number of the virtual machine, name of the machine, a service that runs on this machine, IP address version 4 and version 6, the port where the service is available, the responsible person, notes about the virtual machine, a unique equipment identification number and notification whether the device is automatically monitored or not.

The IP address table contains an unique identification number of IP address, IP version 4, IP version 6, the interface on which IP address is configured and unique equipment identification number. Individual IP address cannot be used as a unique identifier, although it has to be unique for the selected network, because the system can occur two or more same networks and therefore the same IP addresses.

For users system keeps information as follows: user login and password by which the user logs into the system, user role by which we can define user rights in the system.

Registration system is able to provide the following outputs:

- List of devices and their location in laboratories,
- a list of virtual machines that run on registered servers,
- a list of IP addresses that are configured on devices,
- a list of users who have access to the system.

C. *Automatic Detection of Device Availability*

When inserting a new device or virtual machine to the system, Administrator or the Employee may choose whether the system will periodically checked the device or not. This option is enabled only for elements that have their own IP address, ie. working on the third network layer of the OSI / ISO model [6] and support the ICMP protocol [7]. In the case that the automatic availability detection is selected, in the interval of 15 seconds the *Echo Request* message is sent using the ICMP, which verifies whether the device is available and active.

If the device or virtual machine is active, the *Echo Reply* response is coming to the system and the system register the functional element.

However, if the device does not answer for the request within five seconds or come any other ICMP message system re-sent *Echo Request* message. Once, if no answer comes in 5 seconds, the device in the system is marked as unavailable and responsible person is notified by e-mail about problems that arose.

The above described method of detection is shown in Figure 2.

In certain situations, our proposed detection may not function correctly. Some network elements have ICMP disabled for security reasons even if the device is plugged in and fully functional, system those elements marked as unavailable. Likewise, the devices will be marked as unavailable if are located behind NAT.

Therefore, the allowing of the automatic detection is on user consideration, when he/she adding or editing the network device.

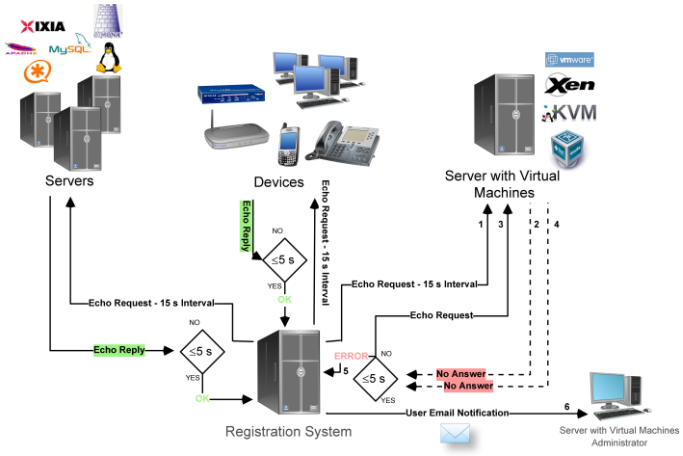


Figure 2. Automatic Detection of Device Availability

D. *Used Technologies*

For the implementation of system ASP.NET 4.0 [8] framework using C# programming language [9] was chosen. One of the main reasons of using a combination of these technologies is good compatibility, speed (response) and also easy implementation. C# language was used because of its extensive capabilities, simplicity, strong support of the various libraries, detailed documentation and many others. ASP.NET framework allows flexible transition from programming traditional desktop tools into web applications. The MySQL database was chosen for data management. This DBMS (Database Management System) is available as free software under the GNU [10]. Another used technology is AJAX [11], which is suitable for the convenience of users. Of course, the CSS [12] styles are used to separate content from design. The next chapter describes the implementation of registration system in details.

III. The Implementation of Registration System and Web Interface

Registration system meets the requirements for separation of application layers (separation of presentation from logic and data). The system is divided into two main sections:

- Data section
- Presentation section

A. *Data Section*

In registration system this part is realized by the MySQL database. Tables are created with tool from Oracle company - SQLdeveloper [13]. Example of SQL script creating table DEVICE:

```
CREATE TABLE Device(

  id_device INT   NOT NULL   AUTO_INCREMENT PRIMARY
KEY,
  type VARCHAR (20)  NOT NULL,
  ip_address VARCHAR (30),
  ip_v6 VARCHAR (50),
  num_u VARCHAR (20),
  maker VARCHAR (20),
  service VARCHAR (20),
  res_person VARCHAR (100),
  name VARCHAR (20),
  building VARCHAR (10)  NOT NULL,
  room VARCHAR (10)  NOT NULL,
  rack VARCHAR (15)
  position VARCHAR(10),
  notes VARCHAR (1000),
  lastChange DATETIME);
```

The script creates the database table named DEVICE, primary key attribute is ID_DEVICE, which is automatically incremented - this means that whenever a new device is added into the DBMS table it finds the biggest used ID_DEVICE in the table, increasing it by one and assign it to a new record.

B. *Presentation Section*

This part of the registration system is implemented like a web application. It consists of separate ASPX [8] files with XHTML [8] codes. For creating user interface an ASP.NET templates called *MasterPages* were used.

Validation of data entry is realized by built-in framework elements like a *RegularExpressionValidator* and *RangeValidator*. System access page is *login.aspx* and each user who wants to work with the system must login and verify his/her identity against the internal account database, or against a remote authentication system.

Upon successful registration system passes to the main application page that contains the menu items depending on the role of the logged user. Menu items are as follows:

Devices - after click to select, the system redirects you to the main page of the system - *index.aspx*. Sub-menu *New device* is available only for users with the role Administrator and Employee and is used to add new devices to the system. Fields marked with a red asterisk must be filled before submitting.

Users - this item is available only for user with the role Administrator, who here can add, edit and delete other users and their roles.

Menu *All devices* is available regardless of user role. It contains *Find device* filter, which is off by default and a table of all devices that are registered in the system. Shown table has the same appearance as the table on the main page. Filter *Find device* can be activated by simply clicking on the *Select* radio button and select the appropriate filter for a list of registered devices. *All Devices* menu contains more sub-menus - *All Virtual Machines* and *All IP Addresses* which provides lists of all registered virtual machines and IP addresses.

The last item is a menu *Change My Password*, which is available regardless of user role. This submenu allows users to change the password for a user account if authentication is realized against the local database.

Further, on the home page of the system a *FindRack* panel and its table are shown. It contains the elements in the selected rack. The *FindRack* panel includes three DropDownLists and button *Find*.

With the first DropDownList (*ddlBuilding*) user selects building, with second (*ddlRoom*) selects a laboratory that is located in the building and the third DropDownList (*ddlRack*) provides selection of a specific rack in laboratory. Choice of options in DropDownList is updated after each selection – if the building is changed, other two DropDownLists (laboratory, rack) are updated and actual options to choose are loaded. This is dealt with using AJAX, which means that the entire page does not reloading, but only values in DropDownLists does. Click on *Find* button cause confirmation to load the selected rack from the database. The rack content is then shown in the table of devices (*DataGridView*) located below the panel *FindRack*. Devices are listed by position in the rack from smallest to largest. This can be easily changed by clicking on the column header by which the user wants to sort the elements. But not all of the devices are shown in the table. When user moves mouse cursor on a device (table row), next to cursor a bubble pops-up with a detailed description of the device. This is realized by using Javascript [14] and AJAX. Home page example screenshot is in the Figure 3.

Figure 3. Registration System

On the left edge of the table, which can be seen in Figure 3, is column named *IP address / Virtual Machine*, which is used for entering IP addresses and edit the parameters for virtual machines. Menu for configuring the parameters of the virtual machines is available only for server-type devices. This menu contains table with virtual machines running on a particular element or the IP addresses assigned to device. This page contains also the button *Add new machine* / *Add new IP*, after click the form for adding a new virtual machine / IP address is loaded.

Furthermore, the table continues with columns *Edit* and *Delete*, which means we can manage the selected device. These columns are displayed only if the logged user has the role Employee or Administrator.

Other column *Type* defines what type of device is registered. After click system goes to a page with details about the selected device – user can find here all information like a virtual machines table, if it is a server, a table of IP addresses and also the parameters provided by the pop-up bubble. Other columns in the table are features of the device listed as follows: position in the rack, the device name, IP address, service, vendor, responsible person and date and time of last modification. Other information about the device, as mentioned above, are available in the *Type* menu or by AJAX bubble after you place your cursor on the table row.

IV. Conclusion and Future Work

The aim of our work was to develop a functional registration system, which graphically describes the various elements of network infrastructure and can also automatically monitor the availability of registered devices. The system can be used not only in laboratories of IP Telephony for which the system is primarily intended, but everywhere the similar issue with managing network infrastructure elements occurs.

A possible extension of the system for the future is to add language switches, solve the problem with ICMP permeability and link the system with the register of buildings and rooms in our university.

The system can be a great benefit for all users, who are looking for tool to simplify their work with network infrastructures used not only for teaching or solving research projects.

Acknowledgment

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 218086.

References

[1] M. Krcmarik, “Evidenční systém aktivních síťových prvků v laboratoři IP telefonie – Bakalářská práce”, Vysoká škola báňská, Technická univerzita Ostrava, 2011.

[2] P. DuBois, “MySQL Cookbook”, O'Reilly Media; Second edition, 976 p., 2007, ISBN-13: 978-0596527082.

[3] G. Carter, “LDAP System Administration”, O'Reilly Media, 312 p., 2003, ISBN-13: 978-1565924918.

[4] J. Hassel, “Radius”, O'Reilly Media, 206 p., 2002, ISBN-13: 978-0596003227.

[5] R. J. Anderson, “Security Engineering: A Guide to Building Dependable Distributed Systems”, Wiley, 1080 p., 2008, ISBN-13: 978-0470068526.

[6] D. Wetteroth, “OSI Reference Model for Telecommunications”, McGraw-Hill Professional Publishing, 320 p., 2001, ISBN-13: 978-0071380416.

[7] J. Postel, “Internet Control Message Protocol”, RFC 792, URL: http://tools.ietf.org/html/rfc792, 1981.

[8] M. MacDonald, A. Freeman, “Pro ASP.NET 4 in C# 2010”, Apress, 1616 p., 2010, ISBN-13: 978-1430225294.

[9] A. Troelsen, “Pro C# 2010 and the .NET 4 Platform”, Apress, 1752 p., 2010, ISBN-13: 978-1430225492.

[10]”GNU General Public Licence”, URL: http://www.gnu.org/licenses/gpl.html, 2007.

[11]N. C. Zakas, J. McPeak, J. Fawcett, “Professional Ajax”, Wrox, 624 p., 2007, ISBN-13: 978-0470109496.

[12]E. A. Meyer, “CSS: The Definitive Guide”, O'Reilly Media, 544 p., 2006, ISBN-13: 978-0596527334.

[13]S. Harper, “Oracle SQL Developer 2.1”, Apress, 496 p., 2009, ISBN-13: 978-1847196262.

[14]D. Flanagan, “JavaScript: The Definitive Guide”, O'Reilly Media, 1100 p., 2011, ISBN-13: 978-0596805524.

Study of VoIP traffic performance under congested MPLS network scenario

Rucka Lukas, Hosek Jiri
Department of Telecommunications
FEEC-Brno University of Technology
Brno, Czech Republic
rucka.lukas@phd.feec.vutbr.cz, hosek@feec.vutbr.cz

***Abstract*—There is a trend that the integrated voice, real-time video and bulk data are together transported through the converged IP/MPLS (Internet Protocol/MultiProtocol Label Switching) core networks. The MPLS technology represents recent advanced way how to optimize and speed up the packet forwarding process in core networks. This paper introduces and evaluates the behaviour of MPLS technology in an experimental IP network with aim to find out the efficiency of the QoS (Quality of Service) mechanisms. The OPNET Modeler simulation environment is used to simulate the network scenario with QoS mechanisms implemented for voice traffic. This paper investigates the performance of the MPLS technology for traffic flows with varying QoS ratings.**

Keywords- MultiProtocol Label Switching; MPLS; OPNET Modeler; VoIP; FEC; Quality of Service

I. INTRODUCTION

Many different types of nowadays applications using the IP (Internet Protocol) networks require more bandwidth, more reliable transfer, low latency, flexibility in the face of failure and faster services. IP networks are datagram packet switched networks that transfer data divided into IP packets. Each IP packet is individually routed through the network based on a hop-by-hop paradigm. When the IP packet arrives to the router, the destination address is examined, router performs a route look-up and send the packet to a matching next hop. This leads to the behaviour, when two packets from the same communication are individually handled in the network. Traditional IP networks offer a little predictability of the service. This behaviour is unacceptable for applications sensitive to a quality of service, such as VoIP (Voice over Internet Protocol), video conference, as well as for all emerging and future real-time applications. Today IP networks transport over the same network many different types of services. Therefore the network should provide a different QoS (Quality of Service) assurance for different types of services [1]. During the past several years, many mechanisms have appeared for providing QoS in IP networks. The fundamental objective of any QoS mechanism in general is to ensure some combination of a high reliability, high bandwidth, low packet drop rate, low delay, low jitter and low packet delay variance. The QoS mechanisms only support prioritization of the traffic and allocation of the capacity under congested conditions, but do not create additional capacity.

Mainly for the backbone network appeared an effort to find a technology or scheme that would reduce the routing load and increase the routing speed. One of the widely accepted results of this effort is a standardized MPLS (MultiProtocol Label Switching) technology [2]. The MPLS technology allows enhancing the performance of the existing protocols over the traditional IPv4 networks and allows the implementation of QoS in the network [3]. The main idea of the MPLS technology is to attach a short fixed-length identifier to packets. This identifier is called label in the MPLS terminology. The MPLS technology can be described as the technology that brought the connection-oriented network to the connection less IP network. Therefore network applications and services can obtain all of the advantages of the IP while they operate over the networks that are reliable and predictable.

II. MULTIPROTOCOL LABEL SWITCHING TECHNOLOGY

The MPLS technology allows the data forwarding through label switching at the second layer (link layer) of the OSI (Open System Interconnection) layer model and integrates routing operations on the third layer of the OSI layer model [2]. A major difference between the MPLS technology and the traditional IP network is in the forwarding technique. The MPLS labels are used for the packet forwarding decision instead of the long and variable IP network addresses. This solution leads to the simplification and acceleration of the look-up and forwarding operations.

A. MPLS Architecture

The part of the network which uses the MPLS technology is often referred to the MPLS domain. The MPLS domain consists of a LSR (Label Switching Router) router whose interfaces operate in the MPLS domain. The LSR router forwards the conventional IP packets as well as the MPLS labelled packets. The LSR router placed at the edge of the MPLS domain is called a LER (Label Edge Router) router. It connects the MPLS domain and the non-MPLS domain (e.g. IP network) together. The flow of the packets incoming from the non-MPLS domain is firstly assigned by labels at the ingress LER router [2]. After that, the packets are forwarded along a LSP (Label Switched Path) path. The LSP path is the one-way routing path operating in the MPLS domain. The LSP path can be established through a LDP (Label Distribution Protocol) [4].

The LDP protocol defines a set of procedures and messages on how one LSR can inform the other LSR on the label bind. The LSP path starts at the ingress LSR router and ends at the egress LSR router. It means that it passes through the MPLS domain. The MPLS technology uses two methods for choosing the LSP path for a FEC (Forwarding Equivalence Class). These methods are called route selection. First method is a hop-by-hop routing and the second method is an explicit routing. In the hop-by-hop routing, each LSR chooses the next hop independently and in the explicit routing, the entire LSP path is specified. Along the LSP path, old labels are replaced with new labels at LSR routers on the path. The label replacement is needed when the packet passes through the LSR router because labels are only locally relevant [2]. Therefore the labels are used to find only the next hop. While label replacement is required along the path, the network layer look-up is not necessary at these LSR routers. It is due to the link layer forwarding with labels. At the egress LER router, labels are completely removed and packets are forwarded to other networks.

MPLS labels are also used to identify a FEC for each of traffic flows. The FEC is a group of packets or flows which are forwarded along the LSP path and are treated in the same way with required forwarding treatment. The MPLS technology performs a packet classification into the relevant FEC class only once. This FEC classification is done by the LER on the edge of MPLS domain. When a packet enters from non-MPLS network into the MPLS domain, the LER decides which packets belong to which FEC. This decision is based on NHLFE (Next Hop Label Forwarding Entry) information and determines where to forward the packet and assigns a label to the packet. This operation is often called label push. Labelled packet is then forwarded towards the subsequent LSR. The NHLFE information consists of the packets next hop and the operation to be performed on the packets label. The incoming label maps each label to the set of NHLFE. NHLFE is used to forward labelled packets as well as unlabelled packets.

B. MPLS and Quality of Service Assurance

It was mentioned above that the label is presented in a header of packet called the Shim Header. The Shim Header is placed between the second layer header and the third layer header. It is 32 bit long and consists of a Label value (20 bits), TC (Traffic Class) field (3 bits), S (Stacking) bit (1 bit) and TTL (Time to Live) field (8 bits). The TC field was initially reserved for an experimental use. From the QoS point of view the TC field is especially important because in the most nowadays MPLS applications it is used as indicator of the QoS requirements.

The IEEE 802.1p recommendation introduced the expression of a CoS (Class of Service) using three precedence bits at the Media Access Control (MAC) layer [5]. These bits can be set for any of eight priorities. The TC field specification in the MPLS technology works in a similar way but it is implemented on the third layer of the OSI layer model. A flexible solution to support DiffServ (Differentiated Services) over MPLS networks was defined in the RFC 5462 [6], [7]. This solution for supporting the DiffServ BA (Behaviour Aggregate) whose corresponding PHB (Per Hop Behaviour) is currently defined over a MPLS network [8], [9], [10]. If in the TC field stored bits are used for QoS classification, then the LSP path is called an E-LSP (Exp-Inferred-PHB Scheduling Class LSP) path [6]. The E-LSP path uses a label as the indication of the FEC destination and TC field is used as the indication of the class of a flow in order to select its PHB including PSC (PHB Scheduling Class) and drop preference. It is important to realize that the DiffServ technology uses six bits to define BA and the corresponding PHB, whereas E-LSP path has only three bits for this function. The L-LSP (Label-Only-Inferred-PSC LSP) path carries packets belonging to one PSC identified by the label value. The PSC consists of one or more PHB where PHB within the PSC are differentiated by TC field value in the Shim Header. The PSC of an L-LSP path is explicitly signalled during the LSP path establishment. In the L-LSP path, the PSC determines the corresponding traffic class and the TC field determines the drop precedence [2].

Each type of LSP path has its advantages and disadvantages. The E-LSP path is easier to operate and more scalable. The E-LSP path preserves labels and uses the TC field for QoS features. But the MPLS technology signals the bandwidth reservation on the per LSP path basis, the bandwidth is reserved for the whole LSP path without the PSC-based granularity. That can lead to lacking of bandwidth in queues serving some particular PSC. The L-LSP path is clumsier, because of more labels for tagging all PSC of all FEC. But if the L-LSP path is used, the label carries the scheduling information and when bandwidth is reserved for a given L-LSP path, it is associated with the priority queue to which this LSP path belongs.

III. MODELLING OF MPLS TECHNOLOGY

The modelling of the MPLS technology was realized in the OM (OPNET Modeler) simulation environment (version 14.5). The basic OM environment has been extended with the MPLS module, which brings simulation capabilities of the MPLS technology. The OM simulation environment provides complex development and simulation environment supporting design, modelling, simulation and analysis of different type of network technologies, architectures and protocols [11]. Both behaviour and performance of the modelled networks and systems can be analyzed by discrete event simulation. The OM also provides tools for the data analysis and presentation. The OM extensively uses a graphical editors to create network topology and for entering the network and model details. These editors provide an intuitive mapping from the modelled network to the OM model specifications.

A. Simulation topology

To evaluate the behaviour of the MPLS mechanism with QoS assurance a large simulation model has been built in the OM. This MPLS model consists of twenty three LER routers and twenty three LSR routers. Sixty eight subnets were connected into the MPLS network. Behaviour of the QoS assurance in the MPLS domain was examined in a part of the large simulation model. Simplified simulation model of MPLS domain is shown in Fig. 1. This topology of MPLS network is composed of two edge LER routers and three LSR routers. MPLS capable routers are connected with E3 Full-Duplex

links. These links are able to transmit data at rates up to 34,386 Mbps.

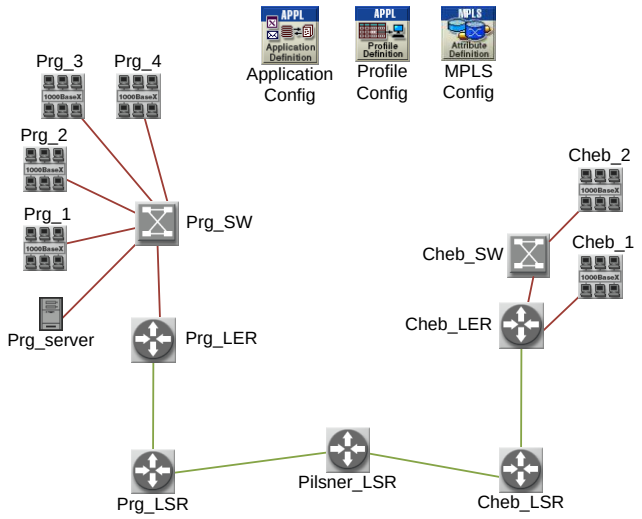


Figure 1. Simplified simulation model of MPLS domain

Other links have higher capacity to minimize the possibility of congestions and make the MPLS domain the bottleneck of the network. It is desired if we want to examine the QoS assurance in the MPLS. The network model also contains three configuration modules, switches, application server and some network subnets. Each subnet is composed of twenty fifth client stations that generate network traffic.

There are two basic components – Application Config and Profile Config which are used for selection and configuration of network applications in the OM. The MPLS Config component is used for a global configuration of the MPLS technology. The CBQ (Class-Based Queueing) queues with DropTail profile are used in the MPLS domain and simple DropTail queues are used in the rest of network. The background network traffic, originated in the subnets and terminated in the Prg_server, has been defined. Its overall speed was 12000 Kbps from each subnet. The MPLS technology was configured on the MPLS routers (LER and LSR). The packets incoming into the ingress LER router were classified into appropriate FEC class based on their application type. The E-LSP dynamic paths were used for connection between LER routers.

B. Simulation scenarios

As mentioned above, the background network traffic has been defined. For more accurate results the additional network application VoIP (Voice over Internet Protocol), FTP (File Transfer Protocol), HTTP (HyperText Transfer Protocol), File Print and Database Access between nodes Cheb and Prague were defined in the simulation model. The additional traffic is originated in Cheb subnets, passes through the Pilsner LSR and destined for the Prg_server. Each traffic type has different demands on QoS. The VoIP traffic has the highest demands on QoS assurance. For this reason, a FEC assigned to VoIP traffic is associated to the second highest priority in the test scenario. The background network traffic and the other data traffic (FTP, HTTP, File Print, Database Access) are associated to the lowest priority. A mixed application traffic profiles are derived from an OM standard application models to represent more realistic network traffic. Simulation models assume the basic characteristics of the mixed application shown in Table I. The data traffic sources are modelled as an exponential traffic sources with a burst and idle periods. By adjusting the duration and repeatability of each application in a traffic profile, different average traffic throughput can be achieved. The traffic generation parameters were chosen in order to achieve full utilization of E3 links in the MPLS domain. Under these conditions, QoS implementation can take a place.

TABLE I. CHARACTERISTICS OF MIXED APPLICATION TRAFFIC PROFILE

Application type	**Parametrs**	
	Interarrival time distribution [s]	***Average object size [kB]***
FTP	exponential (360)	500
HTTP	exponential (60)	875
File Print	exponential (360)	4500
DB Access	exponential (12)	33
VoIP	GSM FR Quality Speech – 13.2 Kbps	

We run two different simulation scenarios for our analysis. In the first case, only the mutual influence of different types of applications is examined in the MPLS domain. In this case, there is no QoS assurance model implemented at all. Thus all traffic flows have to compete on equal basis for the all available network resources. The second scenario takes into account the benefits of MPLS QoS assurance. The VoIP traffic is prioritized over other traffic flows.

The duration of both simulation scenarios is 3600 seconds. Within the period of 0 to 100 seconds only routing information are exchanged through the network. After 100 seconds the VoIP traffic from Cheb_1 subnet to Prg_server become active. 200 seconds from the simulation start-up, the background network traffic is transmitted from all subnets to the Prg_server. Also the additional data traffic (FTP, HTTP, File Print and Database Access) from Cheb_2 subnet to Prg_server starts at this time. In the simulation time of 2880 seconds, all background network traffic ends. VoIP and additional data traffic from Cheb subnets run till the end of simulation. Due to high amount of background traffic transmitted, all E3 Full-Duplex links are fully utilized between time 200 and 2880 seconds in both scenarios.

The characteristics of end-to-end delay, packet delay variation (PDV) and jitter for VoIP traffic were observed. The OM calculates packet delay variation (1) as:

$$PDV = (t_4 - t_2) - (t_3 - t_1). \quad (1)$$

Equation (2) is a jitter calculation in the OM,

$$jitter = (t_4 - t_3) - (t_2 - t_1), \quad (2)$$

where t_1 is the time of sending the first packet, t_2 is the time of sending the second packet, t_3 is the time of arriving the first packet and t_4 is the time of arriving the second packet [11].

IV. Simulation Results

As mentioned in the chapter III. A, there is a bottleneck between Cheb_LER and Prg_LER represented by the E3 Full–Duplex link. For this reason, an increasing amount of transmitted traffic affects the performance characteristics of all flows. The results obtained from the simulations for packet delay variation (PDV), end-to-end delay and jitter for VoIP traffic are shown in the Fig. 2 to Fig. 4.

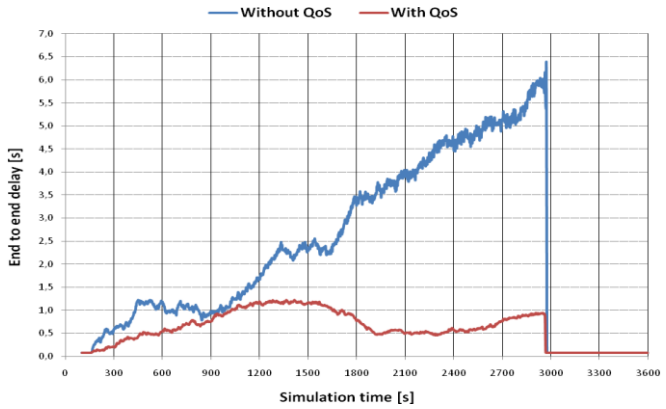


Figure 2. End-to-end delay for VoIP traffic

The Fig. 2 shows the end-to-end delay behaviour of examined VoIP traffic from source node to the destination node. The results for the scenario without providing QoS assurance and the scenario with QoS assurance in the MPLS domain are shown in this figure. It is obvious, that the end-to-end delay in the scenario with QoS assurance shows a much lower values and it is relatively stable throughout the period of high traffic load. In the case of the scenario without QoS assurance, the VoIP end-to-end delay characteristic has ascending tendency. This behaviour is caused by the exhaustion of available bandwidth and by extending of the packets service time in the queues of active network elements. It can be seen that VoIP flow with QoS assurance gives a better end-to-end delay performance. As soon as the period of high traffic load ends, the end-to-end delay goes down to the minimal value immediately for both scenarios.

Next, we examined the packet delay variation (PDV) in both scenarios. In [12] PDV is defined as the difference between two consecutive packets end-to-end delays within a single data stream. The PDV is in the OM calculated according to equation (1). The simulation results of PDV are shown in the Fig. 3. The PDV strongly depends on the end-to-end delay. Therefore, we can see the result trend similar to the Fig. 2. After the beginning of simulation the PDV value increases slowly up to 1500 seconds of the simulation time. From this time until the end of the simulation, the value is stabilized and no longer grows, but only slightly varies. Termination of the background traffic load has no significant effect on the size of PDV value.

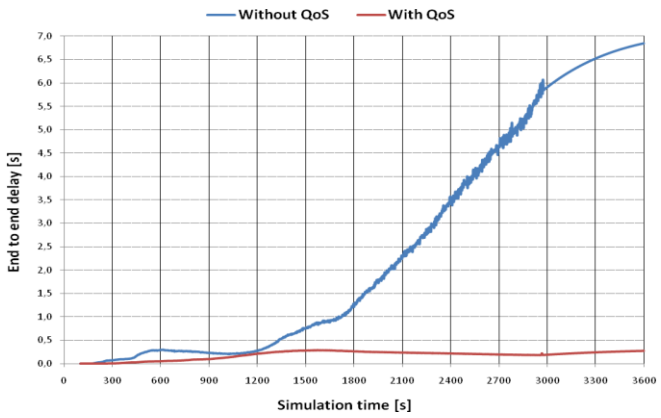


Figure 3. Packet delay variation for VoIP traffic

For the scenario without QoS assurance, we can see relatively low values until 1200 seconds of the simulation time. After this time until the end of the simulation, the PDV value increases continuously. The reason is, as well as in the case of Fig. 2, the exhaustion of the available bandwidth and an increasing queuing time of transmitted packets.

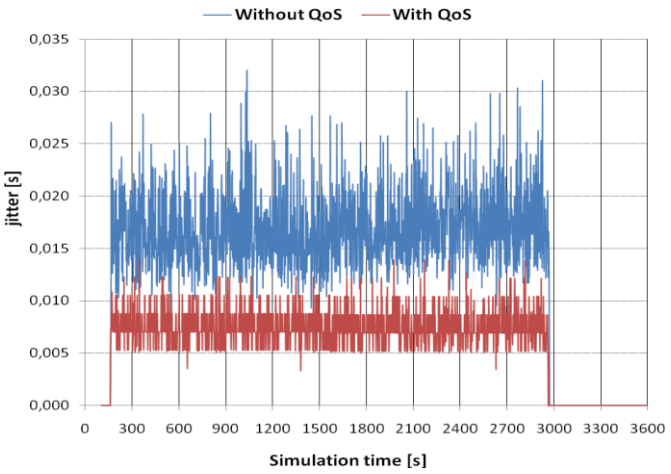


Figure 4. Jitter for VoIP traffic

The Fig. 4 shows the jitter. The jitter of a packet stream is an important QoS characteristic in communication networks. The jitter determines the dimension of the dejitter buffer needed to obtain a regular bit stream at the receiver site. As expected, the result of scenario with QoS assurance gives the narrow jitter with low average value. The result is a consequence of the high priority of the transmitted VoIP traffic. This type of traffic is favoured over others during the active queue management. In the case of scenario without QoS assurance, the jitter is more fluctuating and its average value is higher. The VoIP traffic is classified, as well as other network traffic, into the BE (Best Effort) service class in this scenario. The difference between both scenarios is clearly visible.

V. CONCLUSION

The MPLS technology allows enhancing the performance of the existing protocols over the traditional IPv4 networks. This technology brings the possibility of implementation of transparent QoS schemes inside an MPLS domain. The MPLS technology is composed from several blocks with many control mechanisms. The main goal of this paper was to analyze an impact of QoS assurance in the MPLS technology on the result quality of network applications. The VoIP traffic transmitted over the MPLS domain has been examined and service quality performance has been observed by to the prioritizing of VoIP traffic. For traffic with the highest QoS rating the MPLS technology provides low end-to-end delay, low packet delay variation and narrow jitter with low average value. The network traffic without QoS assurance suffered worse parameters due to links over-utilization.

ACKNOWLEDGMENT

This paper has been supported by the Grant Agency of the Czech Republic (Grant No. GA102/09/1130) and the Ministry of Education of the Czech Republic (Project No. MSM0021630513 and No. 1917/2011/F1/a).

REFERENCES

[1] L. Růčka, J. Hošek, K. Molnár, "Advanced Modelling of DiffServ Technology," In Proceedings of the 32nd International Conference on Telecommunications and Signal Processing - TSP 2009, 2009, pp. 1-6.

[2] E. Rosen, A. Viswanathan, R. Callon, "Multiprotocol Label Switching Architecture," IETF, RFC 3031, January 2001.

[3] J. Hošek, K. Molnár, L. Růčka, D. Komosný, M. Vlček, "Performance Evaluation of Quality of Service Assurance in MPLS Networks," International Journal of Computer Science and Network Security, vol. 2009, no. 10, pp. 114-120.

[4] L. Andersson, I. Minei, B. Thomas, "LDP Specification", IETF, RFC 5036, October 2007.

[5] IEEE 802.1Q-2005, "IEEE standard for local and metropolitan area networks – Virtual bridged local area networks – Amendment 07: Multiple registration protocol", http://www.ieee802.org/1/pages/802.1Q.html, 2005.

[6] F. Le Faucheur, L. Wu, B. Davie, S. Davari, P. Vaananen, R. Krishnan, P. Cheval, J. Heinanen, "Multi-Protocol Label Switching (MPLS) Support of Differentiated Services," IETF, RFC 3270, May 2002.

[7] L. Andersson, R. Asati, "Multiprotocol Label Switching (MPLS) Label Stack Entry: "EXP" Field Renamed to "Traffic Class" Field," IETF, RFC 5462, February 2009.

[8] K. Nichols, S. Blake, F. Baker, D. Black, "Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers," IETF, RFC 2474 (Proposed Standard), December 1998.

[9] J. Heinanen, F. Baker, W. Weiss, J. Wroclawski, "Assured Forwarding PHB Group,", IETF, RFC 2597, June 1999.

[10] K. Molnár, M. Vlček, "Evaluation of Quality-of- Service support in MultiProtocol Label Switching," In Proceedings on the Fifth International Conference on Systems and Networks Communication ICSNC 2010, 2010, pp. 262-265.

[11] OPNET Technologies, OPNET Modeler Product Documentation Release 14.5, 2008.

[12] C. Demichelis, P. Chimento, "IP Packet Delay Variation Metric for IP Performance Metrics (IPPM)," IETF, RFC 3393, November 2002.

IP Telephony in Czech National Research Network

Miroslav Voznak, Filip Rezac, Jan Rozhon, Jiri Vychodil, Karel Tomala
Department of Telecommunications
VSB-Technical University of Ostrava
Ostrava, Czech Republic
miroslav.voznak@vsb.cz, filip.rezac@vsb.cz, jan.rozhon@vsb.cz, jiri.vychodil@vsb.cz, karel.tomala@vsb.cz

Abstract—**The paper deals with the implementation of IP telephony in CESNET and shares the acquired experience as a challenge for designers of communication solutions especially based on open-source platform. IP telephony, as one of the advanced services offered by the Czech Educational and Scientific Network (CESNET), attracts more and more users. For this reason, CESNET puts an appropriate emphasis on further development of the IP telephony infrastructure and improving its quality. About one million voice calls through the CESNET2 network were carried out during 2010. These numbers clearly demonstrate the utility and popularity of this service.**

Keywords- VoIP, IP telephony, CESNET, ENUM, SIP, H.323, Asterisk, Kamailio, OpenSER.

I. INTRODUCTION

The most considerable VoIP implementations at Czech universities are described in this paper. At first, I provide an brief overview of scenarios used. The motivation for deploying each scenario derives from user needs and it is necessary to understand the rationale behind implementing VoIP. I see two basic rationales, the first being an economic impact and the second being an easier integration of information resources into communications. Czech universities apply three different operation modes.

A. PBX's IP trunking

In this mode the existing PBX's of an institution are interconnected through IP (substitution of a simple transmission path with one of very high-level security).

B. IP telephony extensions

The created accounts can be used in SW or HW IP phones (where open-source solution is implemented, IP telephony is strictly based on SIP).

C. SIP trunking

SIP trunking is offered by providers as a service and including multiple voice sessions, about 70 telcommunications companies provide telephony through SIP in Czech Republic. The following chapters describe proposal and advanced implementation of IP telephony which have been arisen in Czech academical environment, concretely in a group of IP telephony acting under Czech Education and Scientific Network association (CESNET).

II. MIGRATION FROM LEGACY PBX TO KAMAILIO PROJECT

The scenario, that is explained in the next subchapters, is implemented at University of West Bohemia (UWB). UWB is a university located in Pilsen in Czech Republic, its IP telephony is based on the Kamailio open-source solution and they are using an interesting autoconfiguration system and self-developed automatic attendant with speech recognition algorithms. Auto-configuration System (AS) enables an automated installation of certain types of Linksys IP phones. It allows multiple phones to be installed without taking up administrators' time usually required to install such phones. The whole AS cooperates with an Kamailio which uses a MySQL database to store its configuration. Once the administrator submits a registration form, the registration systems creates the requested user account in the MySQL database and generates a specific configuration file using a template containing the complete configuration information for an IP phone. The configurations are distributed through TFTP protocol and are downloaded by IP phones when they start up for the first time [1], [2].

A. Migration to open-source IP telephony

This developed solution for IP telephony is based on Linux SIP server with Kamailio (former openSER) and RTP Proxy. Two identical SIP servers in redundancy mode ensure high availability, one is active and the other one in standby, the redundancy feature is controlled by HSRP (Hot Standby Router Protocol). Every server is equipped with two HDD in RAID1 mode, and they are located in separate buildings and designed for 15 000 users. Architecture of the implemented IP telephony at UWB in Pilsen is depicted on Fig. 1. This solution enables a gradual migration from current Siemens hipath 4000 PBX to openSER. The new IP telephony infrastructure with openSER is built as parallel to legacy PBX. Original university telephony network consists of nine PBX Siemens hipath 4000 interconnected through H.323 with central Gatekeeper Siemens hipath 5000. The network management supports not only the mentioned migration from Siemens hipath to OpenSER but also IP telephony provisioning that is described in separate chapter. Telco providers are connected to VoIP Gateways (voip-gw1 and voip-gw2) through ISDN, see the Fig. 1, the main DDI is +420 37763 + four digits extensions. Incoming calls are handled by Voice Gateways (Cisco 2851) and forwarded to the appropriate telephone system, either to Kamailio or Siemens hipath 4000. Outgoing traffic is routed through the Voice Gateways to PSTN. Individual gateways are selected based on the least cost routing principle.

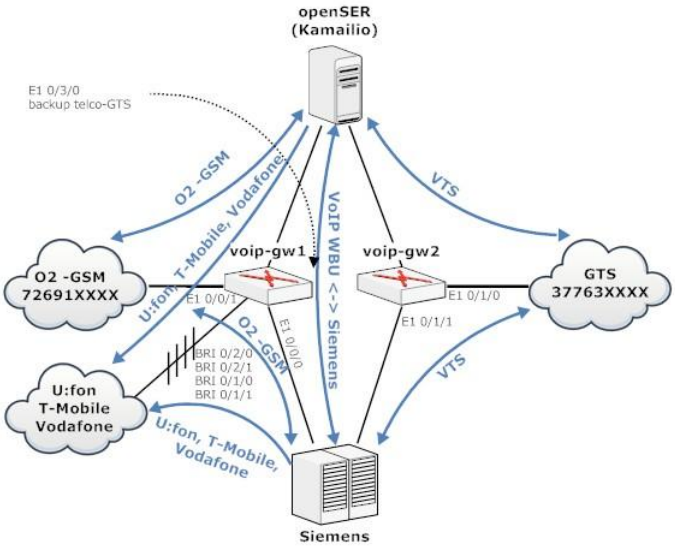


Figure 1. Architecture of the IP telephony at UWB.

SIP server and Voice gateway are the key elements of the presented solution. The extent of the features is generally defined by the OpenSER - Kamalio every configuration of openSER is unique and the system can be customized to fulfil any expectation. OpenSER has recently forked into two projects, Kamailio and OpenSIPS, both solutions encompass the same features so far.

B. IP phones provisioning

The provisioning system at this UWB allows an automated configuration and installation of certain types of 9xx range IP phones by Linksys. The system consists of several parts: LDAP server providing user-specific information used by the Web interface, DHCP server assigning IP addresses dynamically within pre-defined range, TFTP server sharing typical configurations to be downloaded by IP phones when they start up for the first time, and also specific configurations for individual phones, MySQL database storing information on SIP user accounts, Web-based administration interface used to initialize configuration, Request Tracker used to track domain name and IP assignments. The Web administration is user-friendly and Linkys phones offer extended functionality enabling adopt additional information from DHCP server, namely a TFTP server addresses. In the next step the TFTP server provides a provisional initial configuration referring the IP Phone to a specific configuration. The IP phones provisioning system and its interaction with Web based administration is depicted on Fig. 2. The System relies on information entered into the Web Interface consisting of a simple form used to register users and generate configuration files to be stored on the TFTP server. Once the registration form is filled in, the Web Interface also submits a request for an IP address and a domain name to the Request Tracker. Subsequently, appropriate records are created in DHCP and DNS. The IP phone can be connected to the network after the message "registration is successful" was sent to user. MAC addresses have to be specified for hardware phones, SW IP phones do not require that.

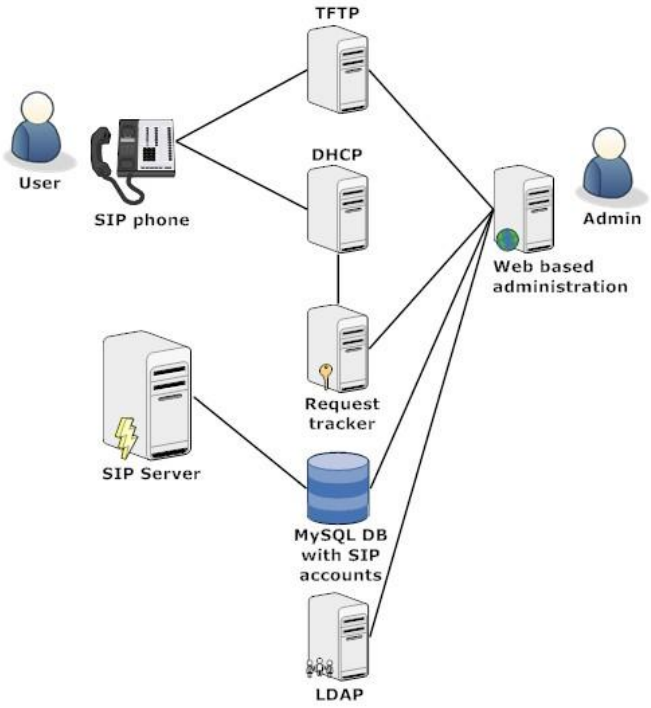


Figure 2. IP phones provisioning system and its interaction with Web based administration.

SIP Passwords are generated automatically since they are only used to configure the IP phones and users do not need to know them. Besides that, the IP phone MAC address and serial number can be now filled in semi-automatically using a bar code scanner. This minimizes the risk of typing errors. All Linksys phones carry appropriate bar codes printed on the outside of their packaging so that it is not even necessary to unwrap them.

C. Automatic Attendant

Department of cybernetics of West Bohemia University applied their own speech recognition algorithms (ASR) to ensure that the called person is recognised and the call transferred to the called party. An automatic attendant allows callers to be automatically transferred to an extension without the intervention of an operator (typically a receptionist). The automatic attendant at this university is a result of long-term research. The first version was developed in 2003. In addition to ASR technology, the automatic attendant involves using dialogue system based on VoiceXML, Oracle database and text to speech (TTS) technology. The SIP interoperability of automatic attendant is ensured by PJSIP open-source client library, the library is multi-platform and enables to include Asterisk in the overall solution. The call flow diagram illustrating the request processing in Automatic attendant is depicted in Fig. 3. Asterisk enables to greet the caller and to replay an announcement. In case the caller is waiting in a call queue, the Asterisk informs him about his position in the queue and finally asks to enter a name of the called person. The task of autoattendant is to analyse the speech data, to look the record up in the database and to ensure that the call is

transferred to the called party. Auto-attendant at West Bohemia University was launched in 2008 and nowadays is able to handle four calls simultaneously.

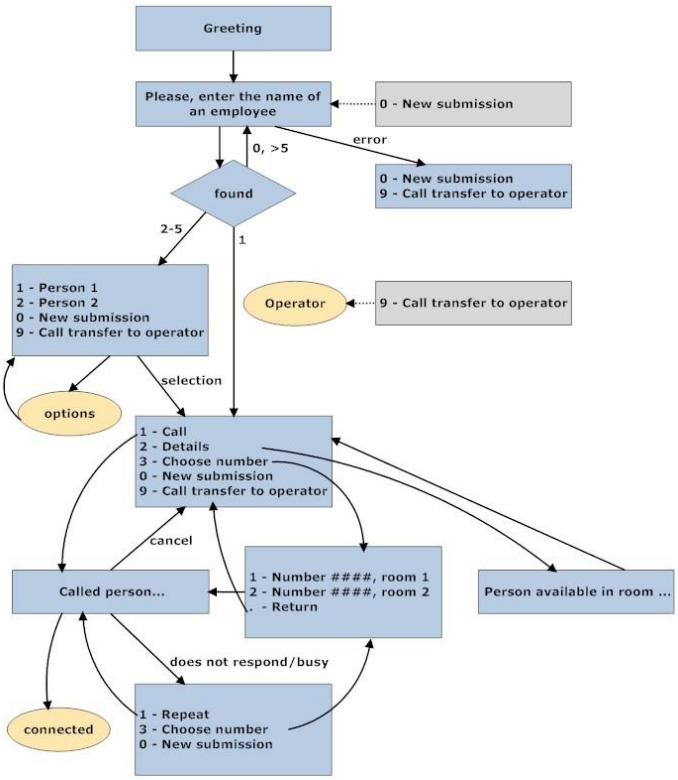


Figure 3. Automatic attendant call flow diagram.

III. INFRASTRUCTURE SOLUTION

In this chapter, two solutions are described, which provide IP telephony to the other academical institutions and thus have to be adapted to the providing of IP telephony services in EDU infrastructure.

A. IP telephony provider in czech EDU

IP telephony provider in czech EDU Czech Technical University in Prague (CTU) has been using a solution based on Cisco Call Manager (CCM) as an extension of current PBX (Ericsson MD110). Unfortunately, this solution is based on the SCCP proprietary protocol defined by Cisco Systems (originally developed by Selsius Corporation). Besides CCM, this university provides voice services for other CESNET members (more than 20) and this solution is based on H.323. This project began ten years ago and within five years nearly all universities had become involved in it. Every CESNET member owns a PBX which can be equipped with a Voice Gateway (VoGW). This VoGW is registered with Gatekeeper and outbound calls to PSTN are routed through VoGW at CTU. CTU makes out the invoices for voice services. The billing system is fed call detail records (CDR's) from every single gateway through RADIUS protocol, CDR's are stored in Postgree SQL database [3].

In 1999, CESNET launched a project offering voice services based on h323 for universities in the Czech Republic. Every member could connect PBX via Voice gateway to the CESNET network and CESNET provided the key elements including gatekeepers. Two gatekeepers ensured the routing between universities and one gatekeeper offered peering to next NREN's and foreign R&E institutions, calls within this infrastructure were free of charge. The infrastructure was gradually extended by additional gateways and in 2001, it was interconnected to a commercial telecommunications operator. The technical solution of the interconnection to a public telephony network required no investments by CESNET2 (connected through the NIX.CZ exchange point). In the same year, the pilot project for calling into the public network was launched and since January 2002, the access to PSTN has been offered as a service to other members involved in the IP telephony project.Nearly all universities joined in this project and about 40 PBX's connected through gateways were registered in 2005. More than 1.5 million voice calls through the CESNET2 network were carried out, with total duration of 4.5 million minutes a year. It was decided to move the paid voice services (peering to PSTN) to another institution because the CESNET's legal status did not allow for providing the services which are commonly available on the market and many IP telephony providers have arisen at that time. Since 2006, Czech Technical University has been providing the voice services with peering to PSTN.

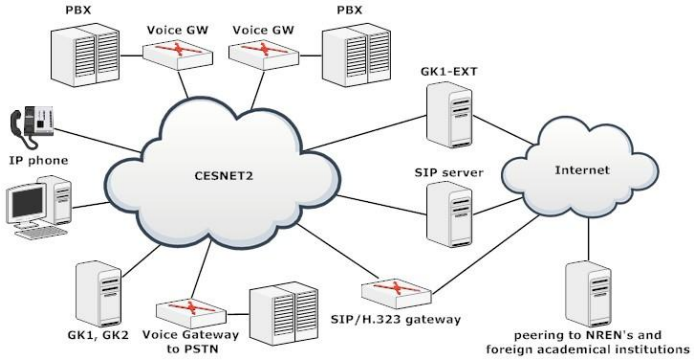


Figure 4. Topology and infrastructure of IP telephony Czech EDU provider.

The infrastructure is natively based on h323 because its original design dates back to 1999. Nowadays, there are gateways without appropriate support widely used SIP protocol. SIP elements have been fully supported by CESNET during the last five years and cooperation with h323 VoIP infrastructure is realised through SIP/H.323 gateways based on Cisco IP2IP located at CESNET. Certain equipment is able to support both protocols, e.g. Asterisk can serves as SIP/H.323 gateway too.

Provision of paid services needs an application enabling administering tariff tables and accounting for calls made by a particular entity. Czech Technical University operates TAS-IP telephony accounting application based on data collected from voice gateways which send information about individual calls through the RADIUS protocol. TAS-IP enables generating both call detail reports and summary reports. Invoices are sent

to individual institutions, the calls performed between universities within the CESNET network are free of charge and calls to PSTN are processed by the TAS-IP billing engine which rates and bills. Through RADIUS, CTU collects not only information for billing but also data about the quality of individual calls. Records are imported into an SQL database which serves as the data resource for own evaluation of the web interface. Cisco gateways evaluates sent the Icpif value calculating estimated speech quality (Icpif means Impairment/Calculated Planning Impairment Factor).

B. SIP Proxy for all and activities of Czech NREN provider

NREN service provider The National Research and Education Network (NREN) providers supply the internet services for research and education communities within a country. CESNET (Czech Education and Science Network) is NREN operator in Czech republic and was held in 1996 by all universities of the Czech Republic and the Czech Academy of Science as an association of legal entities [3]. CESNET has been focusing on IP telephony for a considerable period of time. The activity IP telephony in its research plan was established in 1999, in the first period this activity aimed at implementing H.323, later SIP infrastructure have been built up as a parallel to H.323 with translation gateways based on Cisco IP2IP and Asterisk (oh323 channel). SIP Proxy is the key element of every SIP infrastructure. SIP Proxy at CESNET is powered by SIP Express Router (SER). SER provides functionality of REGISTRAR and PROXY server. SIP clients can register with REGISTRAR and communicate through PROXY, the routing is based on a dedicated number prefixes which are assigned to individual institutions within CESNET. We did not compose a new numbering plan but we adopted the well-known public telephone numbering plan ITU-T E.164. Almost every phone at any Czech university is available at the same number both within the CESNET network and through PSTN. Where it is not possible to communicate with a particular Voice gateway on SIP, then the call is routed through SIP/H.323 gateway. SIP Proxy also handles entire incoming SIP traffic and certain outgoing traffic to other SIP domains such as iptel.org or bts.sk.

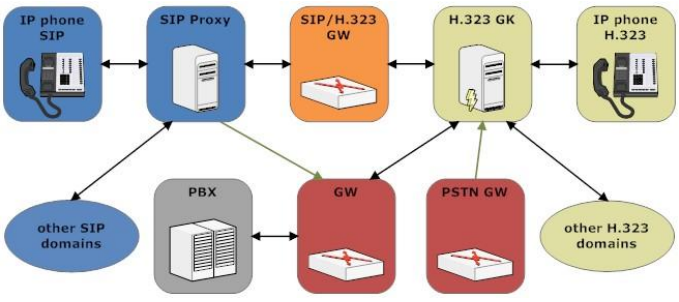


Figure 5. Elements of IP telephony architecture at CESNET.

Nowadays, advanced services have been implemented. Following technical reports regarding CESNET IP telephony were published in the last couple of years.

- IP telephony security overview, CESNET technical report number 35/2006, provides a basic overview of the IP telephony security and focuses in particular on standardised protocols. Its first part explains mechanisms of authentication in protocols SIP and H.323 and the second part deals with attacks, interdomain trust and DNS [4].
- Asterisk and SS7, CESNET technical report number 26/2006, is focused on SS7 networking. Asterisk can interface with both traditional TDM based systems (PSTN networks) and packet based systems (VoIP networks). In our project we focused on interconnecting Asterisk PBX with a PSTN using a Signalling System #7 (SS7) in the role of a call control mechanism [5].
- Open Multiprotocol IP Telephony Dynamic Routing System. The aim of the project, described in CESNET technical report number 20/2006, was to create a multi-protocol system using SIP, H.323 and MGCP standards, which had to ensure routing to various types of VoIP networks. The priority was to provide multi-protocol support to SIP and H.323 signalling and the support of the routing using the ENUM standard (which passed from the trial phase into full operation in the Czech Republic in 2007). The document describes the system's architecture and components used. It also briefly describes ENUM. The appendices list the supported RFC and describe the configuration of individual components [6].
- TLS for SIP Server, CESNET technical report number 13/2007, describes the setup of Transport Layer Security (TLS) in two major open source SIP servers (SER, OpenSer), which are used in the CESNET IP telephony network [7].
- Security Risks in IP Telephony, CESNET technical report 8/2009 deals with VoIP communication security and various techniques of VoIP attacks. We focused our work on Spam over Internet Telephony (SPIT) as a real threat for the future. We have developed both a tool generating SPIT attacks and AntiSPIT tool defending communication systems against SPIT attacks. AntiSPIT represents an effective protection based on statistical blacklist and works without participation of the called party which is a significant advantage [8] and [9].
- SIP Penetration Test System, CESNET technical report 10/2010, deals with a system generating penetration tests, that have to check up the SIP server and test its vulnerability. The tests represent a group at once the most used and effective attacks nowadays, they enable to compose the analysis of security risks and to point out the weaknesses of the tested system [10].

IV. NEW APPROACH TO SUPPLEMENTARY SERVICES

This chapter proposes, specifies and describes implementation of DDDS application for supplementary services provision in IP telephony, this method was developed in CESNET IP telephony group [11]. Supplementary services in traditional telephony were provided by Intelligent Network

(IN) concept developed by ITU. One way to implement the services is by convergence with IN and traditional telephony network. Another more recent activity by IETF is represented by direct implementation of the services by protocol SIP.

Both solutions are either complex and require changes to be made at client/server side or implement only a relatively small subset of services. An alternative method can be the utilization of existing IP mechanisms. An observation has been made that number of IN supplementary services essentially convert identifiers of calling parties. Therefore ENUM (E164 Numbering Maping), as a mechanism that translates telephony and VoIP identifiers, could be used for the service implementation in IP telephony [11].

Because ENUM mechanism is limited in the capability to implement maximum number of IN services, new DDDS application is proposed and specified in this text. ENUM mechanism can be described as an application of Dynamic Delegation Discovery System (DDDS). DDDS application represents an abstract algorithm operating on a database with rewrite rules used by the application for string conversion. In order to design an alternative DDDS application to ENUM several parameters need to be defined: Algorithm, Database and Application specific parameters.

A. Algorithm

General DDDS algorithm was already specified in RFC 3402. The service is provided as a string processing defined by the algorithm, depicted on Fig. 6.

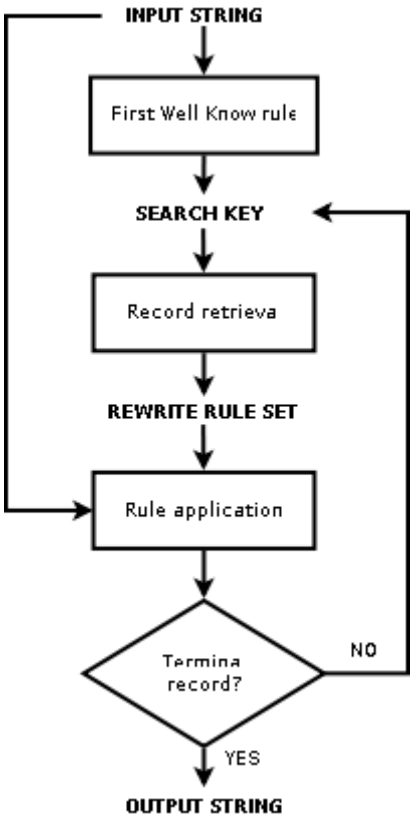


Figure 6. DDDS Algorithm.

The input is initially converted into a database search key used later to query the database. The key is then matched to database records in order to retrieve rewrite rules for input string conversion. In case a rule is not final, it is applied on the initial input and the search is repeated. Once the terminal rule is reached it is applied on the input string to produce output string for further call processing.

B. Database

The database contains rewrite rules for string conversion. DDDS specification does not imply any specific database, however a DNS-based hierarchical system has been proposed in RFC 3403. Properties of well known DNS and its scalability make it desirable storage for application rewrite rules. The rules are stored in format of NAPTR resource records.

C. The Application-Specific Parameters

DDDS application implements the algorithm and makes use of selected database. The application has to specify four parameters: Unique input, First well known rule, Select database and Output. delimiter. The unique input has been defined by ABNF:

uniq-input = fixed-part delim-char dyn-part
delim-char = "&"
fixed-part = uri / e164number
dyn-part = uri / e164number
uri = <URI as defined in RFC 2396>
e164number = <E.164 phone number>

First Well-known Rule (FWKR) is applied on the input and provides a database search key. The specification of the rule conforms to rewrite-rules and is represented in the form of regular expression below. FWKR represents an association between the unique input fixed part and the operator domain for service provider identification. When applied the rule provides domain name search key for DNS query:

!^([^&]*)&!\1.OPERATOR_DOMAIN!

As previously mentioned, proposed DDDS application makes use of DNS to store and query the NAPTR records. The records include regularexpression based rewrite-rule set for string translation. These rules applied to the unique input provide either consecutive search key or final output. The use of the rules represents the service provision in one of three forms: Translation of involved party identifier, Verification of identifier presence, Output selection for routing purpose.

Each rule composes of priority, flags, rewrite rule and service parts. Priority identifies the order in which rules are processed and it can be used for output selection to route the call. Flag indicates, whether a rule is terminal or not and what the following process should be. Rewrite regular expression specifies the translation of string identifier. The regular expression part of NAPTR record for the proposed DDDS application has the ABNF form below. It is composed of two parts. First is regular expression ("ere") that identifies part of the input string intended for subsequent processing. The second part is substitution string ("repl") that contains selected input and an additional part. Application output string is a result of

applying the substitution on unique input string. DDDS application defines two services kinds, where "type" represents a service identifier as specified in ITU recommendation: E2E+type = E.164 numbers translation and E2U+type = E.164 number-to-URI translation.

Output of the application represents the unique input processed by the rewrite rules. Depending on the service type result can be either E.164 number or URI. However some services require an information whether an input string is present in the database instead. Therefore an error channel is added.

The application was implemented for functionality verification in IP telephony environment. Out of numerous IP telephony software, Asterisk PBX was selected as a test platform [11].

V. IP TELEPHONY AS SYSTEM INTEGRATION

In this chapter, a solution at Ostravian University (OU) is described. This university provides IP telephony for their employees with its own developed user-friendly web interface POSERA (PHP OpenSER Administrator). POSERA was implemented in PHP and enables to set up user accounts in OpenSER through the web (HTTPS). The users are verified through LDAP in a corporate directory and then can fill in a form and the new account in OpenSER is created after confirmation. POSERA enables not only creating SIP accounts but also their administration, such as administration of personal information or displaying missed calls. The basic OpenSER configuration was created in user-friendly generator SIP wizard. Compared to the default generated in SIP wizard, some changes to the configuration were made. POSERA is PHP OpenSER Administrator system keeping the same style as the Ostravian University website, depicted on Fig. 7.

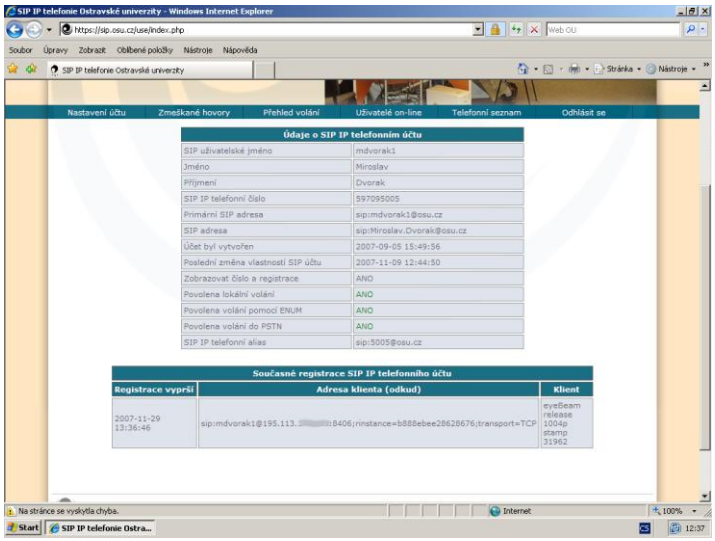


Figure 7. POSERA Web Interface.

Users communicate with the web interface of the SIP server through HTTPS and they are authenticated in LDAP.

VI. CONCLUSION

This paper provides an overview of the significant IP telephony implementations at Czech universities. The advanced best practices are described in individual chapters and all of them are result of research activity in Czech Education and Scientific Network association. Author, as a senior researcher and a founder member of the IP telephony group in CESNET involving staff of Czech universities, decided to describe the most considerable voice over IP implementations at Czech universities and share knowledge with other experts interested in IP telephony.

ACKNOWLEDGMENT

This research has been supported by MSM 6383917201 research intent of the Ministry of Education of the Czech Republic and by the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 218086.

REFERENCES

[1] M. Petrovic, "Manager-Assistant IP Phone Setup," CESNET: Technical report 18/2009, Prague, December 2009.

[2] M. Petrovic, "Linksys SPA9xx IP Phone Autoconfiguration System," CESNET: Technical report 7/2008, Prague, November 2008.

[3] M. Voznak and M. Neuman, "Monitoring and Measurement of Voice Quality in VoIP Environment," CESNET: Technical report 18/2006, Prague, December 2006.

[4] M. Voznak and J. Ruzicka, "IP telephony security overview," CESNET: Technical report 35/2006, Prague, December 2006.

[5] J. Rudinsky, M. Voznak and J. Ruzicka, "Asterisk and SS7, " CESNET: Technical report 26/2006, Prague, November 2006.

[6] M. Voznak, J. Ruzicka and L. Macura, "Open Multiprotocol IP Telephony Dynamic Routing System," CESNET: Technical report 20/2006, Prague, December 2006.

[7] J. Ruzicka, "TLS for SIP Server," CESNET: Technical report 13/2007, Prague, November 2007.

[8] M. Voznak and F. Rezac, "Security Risks in IP Telephony, " CESNET: Technical report 8/2009, Prague, December 2009.

[9] M. Voznak and F. Rezac, "The implementation of SPAM over Internet telephony and a defence against this attack." Publisher: Asszisztencia Szervező Kft. Budapest, 32nd International Conference TSP, Dunakiliti, Hungary, August 2009

[10] M. Voznak, F.,Rezac, and K. Tomala, "SIP Penetration Test System, " In Conference Proceedings TSP 2010 , Baden near Vienna, Austria, August 2010, p.504-508.

[11] M. Voznak and J. Rudinsky, Alternative Methods of Intelligent Network Service Implementation in IP Telephony, The 14th WSEAS International Conference on COMMUNICATIONS, Corfu, July 2010, p.204-207.

Speech and Image Processing

chairman:

Jan Skapa

MOTION ESTIMATION and COMPENSATION for VIDEO OBJECTS

Seyit TUNÇ and Hakkı Alparslan ILGIN
Electronics Engineering Department, Ankara University
Tandogan, 06100, Ankara, Turkey
phone: + (90) 544 66 88 987, email: {stunc, ilgin}@eng.ankara.edu.tr

***Abstract*— Motion estimation is an important process of video compression standards. Motion estimation algorithm we proposed is based on ASIFT technique. In our method, foreground objects are segmented to the regions by using features which are obtained on the object. Depending on the number of features in the region the motion model of the region is determined. By determination of the motion models, motion parameters are estimated. In the motion compensation stage, the motion model is applied on the region and if holes are occurred, they are closed by bilinear interpolation. The visual quality obtained by our method is compared with the ones obtained with most common motion estimation methods.**

Motion estimation; motion compensation; video objects; parameter estimation; video coding; object based motion estimation; affine transformation; perspective transformation

I. INTRODUCTION

Reduction of the storage area occupied by a video signal provides a decrease in hardware cost. High speed transfer by compressing video signal at low bit rates also provides a gain from time. In addition to this, a video signal which uses a small storage area and has a high compression ratio is very important for watching pleasure. For this reason, reducing the size of storage areas, providing a high quality watch and transmitting at high speed by compressing at low bit rates are very important for video signals.

Motion estimation is the most important process used in video compression for reducing temporal redundancy of video sequences. It is used in entire coding standards, from the first video coding standards, such as H.261, to state of the art video coding standard H.264.

While in MPEG-1 [1], MPEG-2 [2], MPEG-4 [5], H.261 [3], H.263 [4] and H.264, [6] frame based motion estimation is used, in MPEG-4, motion estimation for arbitrarily shaped objects are also supported [13].

There are two major factors which determine the success of a motion estimation algorithm. First of them is a low computation time because it is important for real-time coding; and the second one is a high estimation ability since it is important for increasing compression ratio. The method which we proposed focuses on high precise estimation.

At the second section, general information for motion estimation used in video coding standards is given. The proposed motion estimation and compensation method is mentioned at the third section. At the fourth section the proposed method and frequently used motion estimation methods are compared. Finally, results are given at the last section.

II. MOTION ESTIMATION USED IN VIDEO CODING

Most of the progress made in video compression is obtained by improvements in the estimation of motion which are used in video compression algorithms [7].

In video compression standards like H.261, only full pixel motion estimation, and in the MPEG-1 and MPEG-2, which is developed later, a half pixel motion model for making better estimations is used [3]. To obtain a more sensitive motion resolution, in MPEG-4, quarter pixel motion estimation is utilized and with a transformation, for warping of wide regions of frame, a global motion estimation is also applied [14]. In H.264, motion prediction from multi reference frames is used. Also unlike the previous video compression standards which supports only one motion, a multi motion support inside the macro block is added [6].

A. Block Based Motion Estimation

Block based motion estimation is a technique to compute motion of rectangular pixel regions between consecutive frames in MPEG-1, MPEG-2, MPEG-4 and in H.264 [15]. In MPEG-1 and MPEG-2, fixed size square blocks at 16x16 pixels, called macro-block, is used. In addition to the blocks, motion vectors can be computed for arbitrarily shaped regions in MPEG-4. Thus, representation of the movement of an object by only one vector is provided. In H.264, the block size is not limited at 16x16 pixels dimensions, and can be at 16x8, 8x16, 8x8, 8x4, 4x8, 4x4 pixels dimensions. In this way, the blocks can align to the regions, which consist of motion discontinuities, like edges of a moving object.

B. Object Based Motion Estimation

Object based video coding used in MPEG-4 provides high coding efficiency as well as opportunity to perform content based applications because it relies on shape information of moving objects [14].

In object based motion estimation, video images are segmented into moving foreground objects rather than splitting

into the blocks. Motion of each object is estimated independently from the others. For this reason, in video coding, object based motion estimation carries out more motion estimation [8, 16].

III. Proposed Motion Estimaton And Compensation Method

In the proposed method, corresponding point pairs are obtained on foreground objects in consecutive frames for motion estimation. Extracted corresponding point pairs are clustered and these clustered pairs generate the seeds of the regions which will be segmented on the objects. With the help of the clustered corresponding point pairs, regions arise on the object. According to the number of corresponding point pairs, the motion transformation of the region which will be applied is determined. When motion transformation for each region is determined, motion estimation for foreground object is obtained.

Estimated motion transformation for each region, is applied to the foreground object at the reference frame and the holes emerged are filled by interpolation. Thereby, compensation of the motion estimation is provided.

A. Segmentation of the Regions

By using ASIFT, [9] on the moving objects at two consecutive frames, corresponding point pairs are extracted figure 1 (a), (b).

After acquiring corresponding point pairs these pairs are clustered by K-Means Clustering [10], as shown in Figure 1 (c). Then by using Region Growing [11], pixels around each point pair are joined to the cluster of those point pairs iteratively. Thus, regions where motion will be estimated on the foreground object are segmented, which is shown Figure 1 (d).

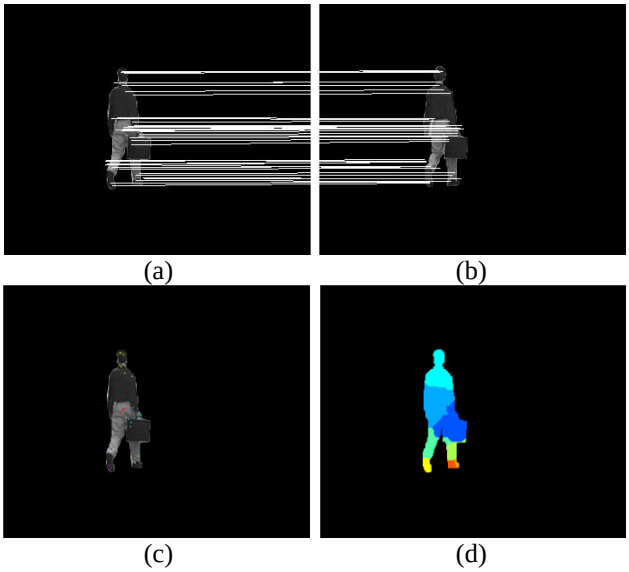


Figure 1. Segmentation of the regions. (a) and (b) are 48th 49th frames of the Hall Monitor. Corresponding point pairs are extracted by using ASIFT. (c) Pairs clustered by k-means clustering. (d) Regions segmented on the object.

B. Motion Models of the Regions

After foreground object is segmented to the regions, depending on the number of corresponding points in the region, the motion model of the region is determined. For example, to find unknowns m and n at the translation motion model in (1), only one corresponding point in the region is enough because every single corresponding point pair ((x, y), (x', y')) gives us two equations.

$$x' = x + m \qquad y' = y + n \tag{1}$$

If another pair exists in the region, anisotropic scaling and translation motion will be as follows

$$x' = ax + m \qquad y' = bx + n \tag{2}$$

or rotation, isotropic scaling and translation motion will be modeled as in (3).

$$x' = a\cos\theta x - a\sin\theta y + m \qquad y' = a\sin\theta x + a\cos\theta y + n \tag{3}$$

If there is also a third corresponding point in the region, affine motion model in (4) can be calculated [17].

$$x' = ax + by + m \qquad y' = cx + dy + n \tag{4}$$

If there is four or more corresponding point in the region, in addition to translation, rotation and scaling, a perspective motion model including perspective is presented in (5) [18].

$$x' = (ax + by + m) / (px + qy + 1) \qquad y' = (cx + dy + n) / (px + qy + 1) \tag{5}$$

C. Motion Estimation

To solve affine motion model in (4), when we rewrite the equation in matrix form

$$\begin{bmatrix} x'_1 & y'_1 \\ x'_2 & y'_2 \\ x'_3 & y'_3 \end{bmatrix} = \begin{bmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \\ m & n \end{bmatrix} \tag{6}$$

is acquired. When a is 1, b is 0, c is 0 and d is 1, only translation motion model is obtained. Translation motion is determined by leaving m and n in (1).

Affine motion estimation is done by taking inverse of the first matrix on the right hand side of the (6). While computing the inverse of the matrix, SVD [12] is used.

To solve motion model in (5), we rewrite for n amount of corresponding point pairs, in matrix form

$$\begin{bmatrix} x_1 & y_1 & 1 & 0 & 0 & 0 & -x'_1 x_1 & -x'_1 y_1 & -x'_1 \\ 0 & 0 & 0 & x_1 & y_1 & 1 & -y'_1 x_1 & -y'_1 y_1 & -y'_1 \\ . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . \\ x_n & y_n & 1 & 0 & 0 & 0 & -x'_n x_n & -x'_n y_n & -x'_n \\ 0 & 0 & 0 & x_n & y_n & 1 & -y'_n x_n & -y'_n y_n & -y'_n \end{bmatrix} \begin{bmatrix} a \\ c \\ m \\ b \\ d \\ n \\ p \\ q \\ r \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ . \\ . \\ 0 \\ 0 \end{bmatrix} \quad (7)$$

is obtained. This equation is solved by homogeneous linear least squares to find transformation coefficients since it is in Ax = 0 fom in general.

D. Motion Compensation

Motion transformation calculated for each region on the foreground object is applied to that region. After applying the transformation, especially in cases where the object gets closer to the camera, holes can occur on the object. In such cases, the holes are filled by bilinear interpolation since its complexity is lower and it gives adequate results.

IV. EXPERIMENTAL RESULTS

The achievement of the proposed method is compared by block based motion estimation methods. These methods are Exhaustive Search (ES), Three Step Search (TSS), New Three Step Search (NTSS) [19], Simple and Efficient Three Step Search (SESTSS) [20], Four Step Search (FSS) [21], Diamond Search (DS) [22] and Adaptive Rood Pattern Search (ARPS) [23].

A car video sequences taken by fixed camera in 640x480 sizes is used to test achievement of the proposed method for translational motion. The proposed motion estimation method and block based motion estimation methods are applied to the car video sequences and objective video quality is acquired by PSNR (Peak Signal to Noise Ratio). 16x16 pixels block size and +/-7x7 pixels search region are used for block based motion estimation methods. 8x8 pixels block size is also used for exhaustive search which gives the best result through block based motion estimation methods.

10th and 11th frames of the car video sequences are shown in Figure 2 (a) and (b) respectively. PSNR of the frame, compensated by exhaustive search method fulfilled using 8x8 pixels block size, shown in Figure 2 (c) is 29.6773 dB. PSNR of the frame, compensated by the proposed method, is obtained as 31.0734 dB.

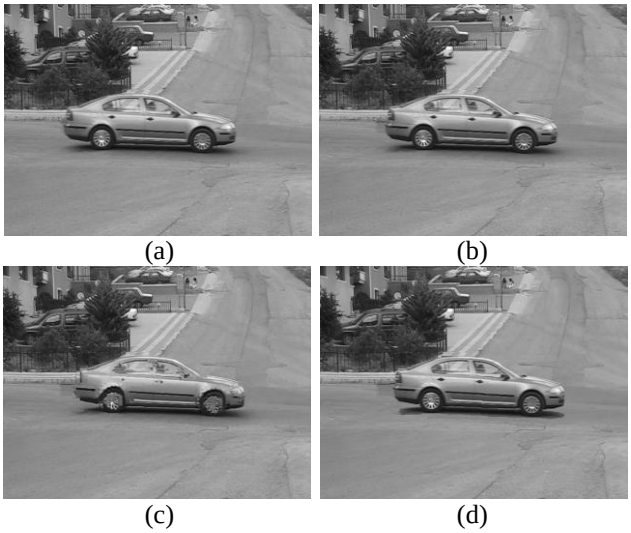


Figure 2. Frames of the translation test video. (a) 10th frame (b) 11th frame (c) The compensated 11th frame by exhaustive search method with 8x8 pixels block size (PSNR=29.6773). (d) The compensated 11th frame by proposed method (PSNR=31.0734).

PSNR results acquired by different methods applied in consecutive 20 frames of car video sequences are shown in Figure 3.

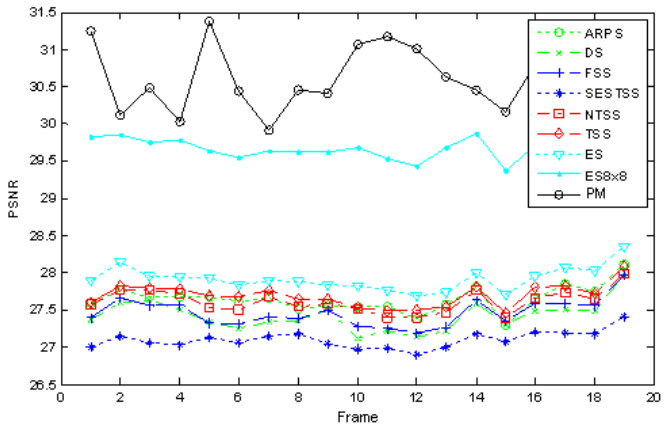


Figure 3. PSNR (dB) values obtained in the translation test video.

Speed of the car are equal between whole test video. Therefore, PSNR values of the block based motion estimation are not hugely different between the frames as shown in Figure 3. It is seen that the proposed method (PM) gives better performance than the other methods. The main reason of this achievement is that motion estimation is independent from speed of object motion in the proposed method.

Average PSNR results obtained for each methods in the car video sequences are shown in Figure 4.

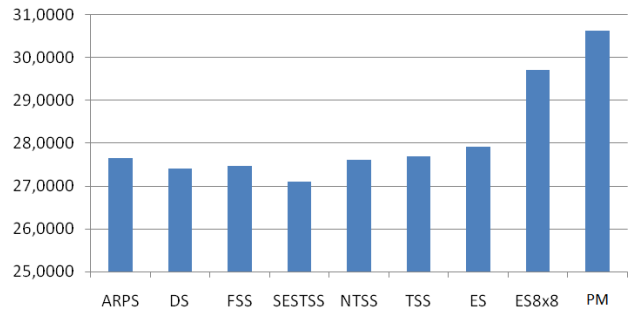


Figure 4. Average PSNR (dB) values for each methods in the translation test.It is seen from Figure 4 that the proposed method (PM) gives better result approximately 3 dB than exhaustive search with 16x16 pixels block size and approximately 1 dB than exhaustive search with 8x8 pixels block size.

A test video is also prepared with an illusion image [24] in 800x592 resolutions to test the proposed motion estimation method's success for rotational movements. In the illusion when the image is rotated by 180o angle, a donkey changes to a beaver. The proposed motion estimation method and block based motion estimation methods are applied to 30 consecutive frames in the prepared video and objective video quality is obtained by PSNR. 16x16 pixels block size and +/-7x7 pixels search region are used for block based motion estimation methods. 8x8 pixels block size is also used for exhaustive search which gives the most successful result through block based motion estimation methods.

15th and 16th frames of the rotation test video are shown in Figure 5 (a) and (b) respectively.

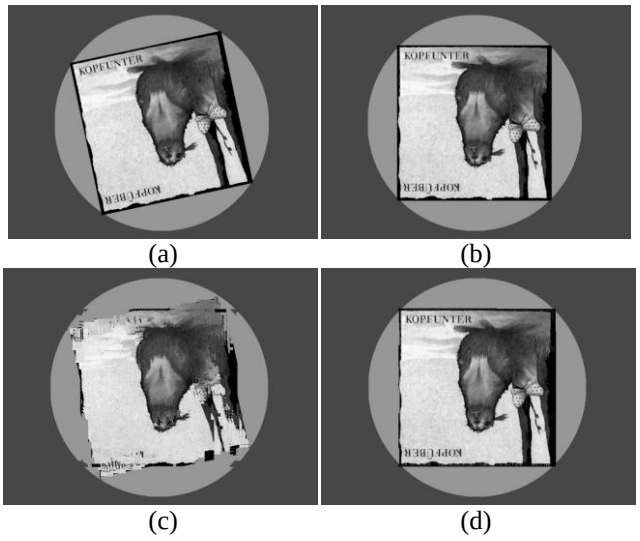


Figure 5. Images of the rotation test video. (a) 15. frame (b) 16. frame (c) The compensated 16. frame by exhaustive search method with 8x8 pixels block size (PSNR=18.1092). (d) The compensated 16. frame by proposed method (PSNR=27.9926).

PSNR results obtained by different methods applied in consecutive 30 frames of the rotation test video are shown in Figure 6.

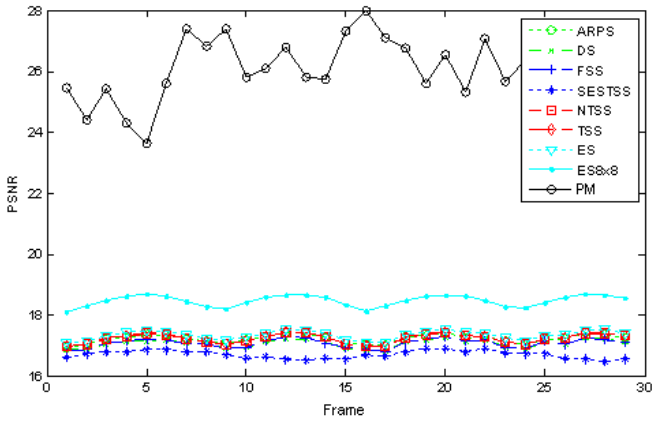


Figure 6. PSNR (dB) values obtained in the rotation test video.

When the results obtained in the rotation test video are examined, it is seen that the proposed method is more successful than the other methods. The main reason of this performance is that object motion is also modelled by affine transformation in the proposed method. On the other hand, only translation motion model is calculated in motion estimation applied in the other methods.

Average PSNR results acquired for each methods in the rotation test are shown in Figure 7.

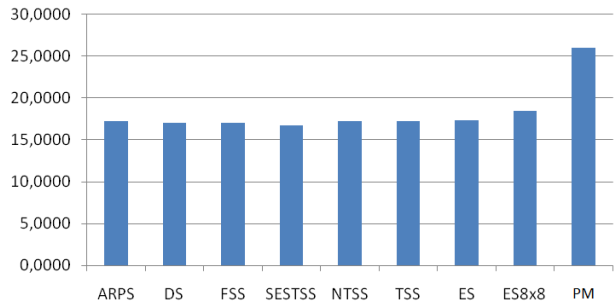


Figure 7. Average PSNR (dB) values for each methods in the rotation test.

It is seen from Figure 7 that the proposed method (PM) gives better result approximately 8.5 dB than exhaustive search with 16x16 pixels block size and approximately 7.5 dB than exhaustive search with 8x8 pixels block size.

V. CONCLUSION

In this study, an object based motion estimation method is proposed for videos taken by fixed cameras. In the proposed method, motion estimation is executed using silhouette of objects, and motion of objects is modelled by scaling, rotation and perspective motion together with translation motion. Estimation is independent from speed and motion model of moving object. As seen from the results the proposed method gives better results than the block based motion estimation methods.

Consequently, our method gives better results for the video sequences which have different motion properties such as

rotation and translation since it considers not only translation information but also rotation, anisotropic scaling, isotropic scaling and perspective.

REFERENCES

[1] ISO/IEC 11172, Information technology – coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s, 1993.

[2] ISO/IEC 13818, Information technology: generic coding of moving pictures and associated audio information, 1995.

[3] ITU-T Recommendation H.261, Video CODEC for audiovisual services at px64 kbit/s, 1993.

[4] ITU-T Recommendation H.263, Video coding for low bit rate communication, Version 2, 1998.

[5] ISO/IEC 14496-2, Coding of audio-visual objects – Part 2: Visual, 2004.

[6] ISO/IEC 14496-10 and ITU-T Rec. H.264, Advanced Video Coding, 2003.

[7] M. Ghanbari, Image Compression to Advanced Video Coding, UK, Hertz, 2003.

[8] M. Phadtare, "Motion Estimation Techniques in Video Processing", Electronic Engineering Times India, 2007.

[9] J. M. Morel and G. Yu, "ASIFT: A New Framework for Fully Affine Invariant Image Comparison", SIAM Journal on Imaging Sciences, vol. 2, issue 2, 2009.

[10] S. P. Lloyd, "Least squares quantization in PCM". IEEE Transactions on Information Theory, 129–137, 1982.

[11] W. K. Pratt, Digital Image Processing 4th Edition, John Wiley & Sons, Inc., Los Altos, California, 2007.

[12] M. R. Hestenes, "Inversion of Matrices by Biorthogonalization and Related Results". Journal of the Society for Industrial and Applied Mathematics, 51–90, 1958.

[13] T. Sikora, "The MPEG-4 Video Standard Verification Model", IEEE Transaction on Circuits and Systems for Video Technology, Vol. 7, No. 1, 1997.

[14] M. C. Lee et al., "A Layered Video Object Coding Sysem Using Sprite and Affine Motion Model", IEEE Transaction on Circuits and Systems for Video Technology, Vol. 7, No. 1, 1997.

[15] I. E. G. Richardson, "H.264 and MPEG-4 Video Compression ", Wiley, 2003.

[16] Y. Yu, D. Doermann, "Model of Object-Based Coding for Surveillance Video", ICASSP, 2005.

[17] G. Forsythe, M. Malcolm, and C. Moler, "Computer methods for mathematical computations, Prentice-Hall, 1977.

[18] R. Hartley, A. Zisserman, "Multiple View Geometry in Computer Vision", Cambridge University Press, 2003.

[19] Li, R., Zeng, B. and Liou, M. L., "A New Three-Step Search Algorithm for Block Motion Estimation", IEEE Trans. Circuits And Systems For Video Technology, vol 4., no. 4, pp. 438-442, 1994.

[20] Lu, J. and Liou, M. L., "A Simple and Efficent Search Algorithm for Block-Matching Motion Estimation", IEEE Trans. Circuits And Systems For Video Technology, vol 7, no. 2, pp. 429-433, 1997.

[21] Po, L. M. and Ma, W. C., "A Novel Four-Step Search Algorithm for Fast Block Motion Estimation", IEEE Trans. Circuits And Systems For Video Technology, vol 6, no. 3, pp. 313-317, 1996.

[22] Zhu, S. and Ma, K. K., "A New Diamond Search Algorithm for Fast Block-Matching Motion Estimation", IEEE Trans. Image Processing, vol 9, no. 2, pp. 287-290, 2000.

[23] Nie, Y. and Ma, K. K., "Adaptive Rood Pattern Search for Fast Block-Matching Motion Estimation", IEEE Trans. Image Processing, vol 11, no. 12, pp. 1442-1448, 2002.

[24] Jusim, J. and Pressler, M., "Kopfunter Kopfüber", Carl Hanser, 1999.

Comparison of MPEG compression standards using different subjective video quality methods

Miroslav Uhrina, Martin Vaculik
Department of Telecommunications and Multimedia
University of Žilina
Žilina
miroslav.uhrina@fel.uniza.sk, martin.vaculik@fel.uniza.sk

***Abstract*—This article focuses on the comparison of the well-known MPEG compression standards (MPEG -2, MPEG-4 Part 2, MPEG-4 Part 10) which were evaluated by different subjective video quality metrics (DSIS, DSCQS, ACR). This testing is the first part of the research of a global video quality factor which will rate not only the impact of a compression on the video quality but also the impact of the transmission link. In the first part short characteristics of the MPEG compression standards are written. In the second part the subjective video quality methods used in our experiments are described. The last part of this article deals with the measurements and experimental results.**

Keywords-MPEG, subjective assessment, DSIS, DSCQS, ACR.

I. INTRODUCTION

In the last years the field of a multimedia technology has rapidly increased. Many new compression techniques and standards have been developed, most of them based on the MPEG technology. The compression is one of the most common factors that influence the video quality. Due to this reason, the video quality assessment as a part of the multimedia technology has become an important role.

II. MPEG COMPRESSION STANDARDS

MPEG, which stands for Moving Picture Experts Group, is the name of a family of standards used for coding audio-visual information (e.g., movies, video, music) in a digital compressed format [1].

MPEG-2 is one of the most used compression standards. It was approved in 1994. MPEG-2 is an extension of MPEG-1 and the video coding scheme is a refinement of MPEG-1 standard. Unlike MPEG-1 MPEG-2 also supports the interlaced video. Furthermore many new functionalities such as scalability were introduced. MPEG-2 also defines the profiles and levels. A profile describes a degree of functionality whereas a level describes resolution and bitrates. But not all levels are supported at all profiles. The most important application of the MPEG-2 standard is TV broadcasting (DVB-T, DVB-S, DVB-C) but it is used also for storage the movies on DVD and other similar disks [2], [3], [4], [5].

MPEG-4 Part 2 (Visual) is the combination of standard coding and object coding. It was approved in 1998 and improves on the popular MPEG-2 standard both in terms of compression efficiency and flexibility. It achieves this in two main ways, by making use of more advanced compression algorithms and by providing an extensive set of "tools" for coding digital media. Some of the key features that distinguish MPEG-4 Visual from previous coding standards include:

- efficient compression of progressive and interlaced video sequences,
- coding of video objects (irregular-shaped regions of a video scene),
- support for effective transmission over networks,
- coding of still "texture" (image data),
- coding of animated visual objects such as 2D and 3D polygonal meshes, animated faces and animated human bodies,
- coding for specialist applications such as "studio" quality video [6].

The latest and currently most used compression standard designed for a wide range of applications, ranging from video for mobile phones through web applications to TV broadcasting (HDTV) is MPEG Part 10 (H.264/AVC). The key advantages of this standard are:

- up to 50% bit rate saving,
- high quality video,
- error resilience,
- network friendliness.

Some of the feature enhancements in MPEG-4 Part 10 (H.264/AVC) standard over the earlier codecs are:

- DCT algorithm works at 4x4 pixels instead of 8x8 but also supports 8x8,
- DCT is layered using Hadamard transforms,
- colour sampling supported at 4:2:2 and 4:4:4.,
- up to 12 bits per pixel are possible,
- motion compensation blocks are variable sizes,
- arithmetic variable-length coding,
- built-in de-blocking filter and hinting mechanism,
- rate-distortion optimizer,
- weighted bi-directional prediction,
- redundant pictures,

- flexible macroblock ordering,
- direct mode for B-frames,
- multiple reference frames,
- sub-pixel motion compensation.

MPEG-4 Part 10 (H.264/AVC) also defines the profiles and levels but its organization is much simpler than in MPEG-4 Part 2. There are only three profiles currently defined (Baseline, Main, Extended) [2], [5], [7], [8].

III. SUBJECTIVE VIDEO QUALITY METHODS

The video quality evaluation can be divided into objective and subjective assessment. The subjective assessment consists of the use of human observers (people) who watch the sequences and score the video quality. It is the most reliable way how to determine the video quality and should not be replaced with objective assessment. The disadvantage of this method is that it is time consuming and human resources are needed. Owing to this fact, the objective methods are mostly preferred and used.

The well-known subjective methods are:

- Double Stimulus Impairment Scale (DSIS) also known as Degradation Category Rating (DCR),
- Double Stimulus Continuous Quality Scale (DSCQS),
- Single Stimulus Continuous Quality Evaluation (SSCQE),
- Absolute Category Rating (ACR) also known as Single Stimulus (SS),
- Simultaneous Double Stimulus for Continuous Evaluation (SDSCE) [2], [9].

To achieve reliable results, minimum 15 observers should be used. They should be non-experts, in the sense that they are not directly concerned with television picture quality as a part of their normal work and they are not experienced assessors. The number of assessors needed for the tests depends upon the sensitivity and reliability of the test procedure adopted and upon the anticipated size of the effect sought. Before the test session assessors should be introduced to:

- the method of a assessment,
- the types of impairments,
- the grading scale,
- the sequence,
- the timing (the reference and the test sequence time duration, the time duration for voting).

The presentation structure of a test session is shown in the figure 1 [2], [9], [10].

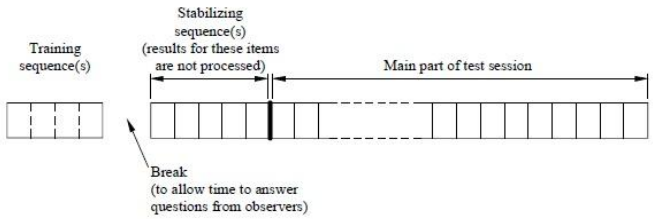


Figure 1. The presentation structure of the test session.

The source signal provides the reference sequence directly and the input for the system under test. It should be of optimum quality (without failing). To obtain stable results the absence of defects in the reference sequence is crucial [2], [9], [10].

The training sequences demonstrate the range and the type of the impairments that are assessed. They should be used with illustrating pictures other than those used in the test but of comparable sensitivity [2], [9], [10].

The whole session should last up to 30 minutes. At the beginning of the first session some sequences (from three to five) should be shown to stabilize the observers´ opinion. The data obtained from these presentations must not be taken into account in the results of the test. A random order should be used for the presentations but the test condition order should be arranged so that any effects on the grading of tiredness or adaptation are balanced out from session to session. Some of the presentations can be repeated from session to session to check coherence [2], [9], [10].

Finally after the test session the calculation of the mean score is done:

$$\overline{u_{jkr}} = \frac{1}{N}\sum_{i=1}^{N} u_{ijkr} \qquad (1)$$

where: u_{ijkrs}: score of observer i for test condition j, sequence k, repetition r,

N: number of observers [2], [9], [10].

In our experiments DSIS, DSCQS and ACR methods were used.

A. *The Double-Stimulus Impairment Scale Method (DSIS)*

In this method two sequences are shown to the assessor - the unimpaired (the reference) sequence and the same sequence impaired (the test one). The reference sequence is shown before the test one (figure 2) and the viewer knows which one is the reference and which one the test.

Reference sequence	Test sequence	Vote

Figure 2. The presentation structure of the test material

After watching both sequences the assessor is asked to rate the second, keeping in the mind the first. The five-grade impairment scale is used:

- 5 imperceptible,
- 4 perceptible, but not annoying,

- 3 slightly annoying,
- 2 annoying,
- 1 very annoying [2], [9], [10], [11].

B. *The Double-Stimulus Continuous quality-Scale Method (DSCQS)*

By this method also two sequences are shown to the assessor - the unimpaired (the reference) sequence and the same sequence impaired (the test one) but the viewer is not informed which one is the reference and which one is the test (figure 3). The position of the reference sequence is changed in pseudo-random fashion.

A	B	A	B	Vote

Figure 3. The presentation structure of the test material

The participant sees sequence A then sequence B twice and then is asked to rate both sequences. The grading scale is divided into five equal intervals:

- excellent (80 - 100),
- good (60 - 79),
- fair (40 - 59),
- poor (20 - 39),
- bad (10 - 19) [2], [9], [10], [11].

C. *The Absolute Category Rating Method (ACR)*

This method is also called single stimulus method (SS). In this method only the impaired (the test) sequence is shown to the assessor (figure 4) so the viewer does not know which quality is the reference sequence.

Test sequence	Vote

Figure 4. The presentation structure of the test material

The assessor is asked to rate the quality of the test sequence based on the level of the quality he has in his opinion for it after watching it. The five-level grading scale is used:

- 5 excellent,
- 4 good,
- 3 fair,
- 2 poor,
- 1 bad [2], [9], [10], [11], [12].

IV. MEASUREMENTS

In our experiments two test sequences were used – one with slow motion (called the "Train" sequence) and one with dynamics scene (called the "Football" sequence). Both sequences have the resolution 720x576 px and 25 fps (frames per second). The length of these sequences were 220 frames, i.e. 8,8 seconds. Both sequences were downloaded in the uncompressed format (*.yuv) from [13] and used as the reference sequences. Afterwards they were encoded to different MPEG compression standards (MPEG-2, MPEG-4 Part 2, MPEG-4 Part 10) using the FFMPEG program version SVN-r24872. The parameters of the encoded sequences were set to Main Profile, Main Level for MPEG-2 standard; Main Profile, Level 3 for MPEG-4 Part 2 standard; Main Profile, Level 3.1 for MPEG-4 Part 10 standard. The target bitrates were in range from 3 Mbps to 15 Mbps, changed in 2 Mbps step. For the subjective evaluation DSIS, DSCQS and ACR methods were chosen.

By the DSIS method 19 assessors were used (16 men and 3 women). Their age was in range from 20 to 71 years, the mean age was 27 years.

The figure 5 shows the results of the performed tests of the "Train" sequence using DSIS method and the figure 6 shows the results of the performed tests of the "Football" sequence using DSIS method.

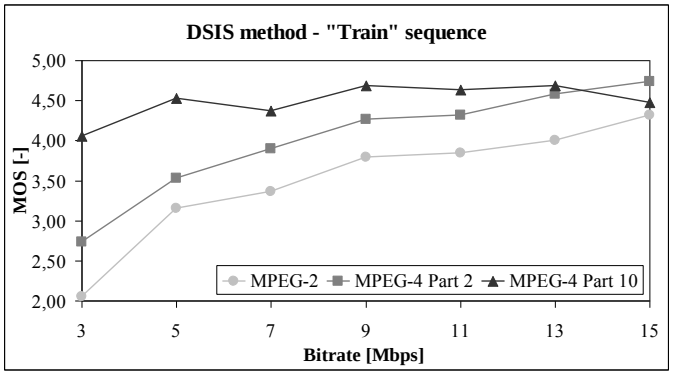


Figure 5. The results of the "Train" sequence using DSIS method

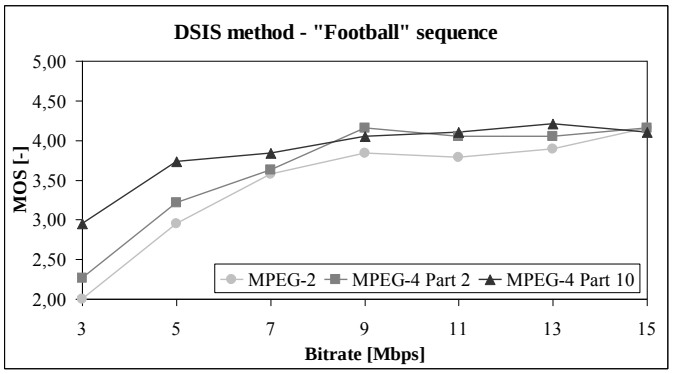


Figure 6. The results of the "Football" sequence using DSIS method

By the DSCQS method 17 assessors were used (15 men and 2 women). Their age was in range from 20 to 23 years, the mean age was 21 years.

The figure 7 shows the results of the performed tests of the "Train" sequence using DSCQS method and the figure 8 shows the results of the performed tests of the "Football" sequence using DSCQS method.

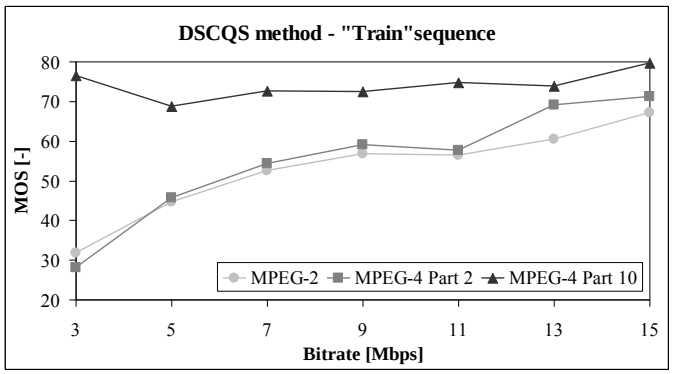


Figure 7. The results of the "Train" sequence using DSCQS method

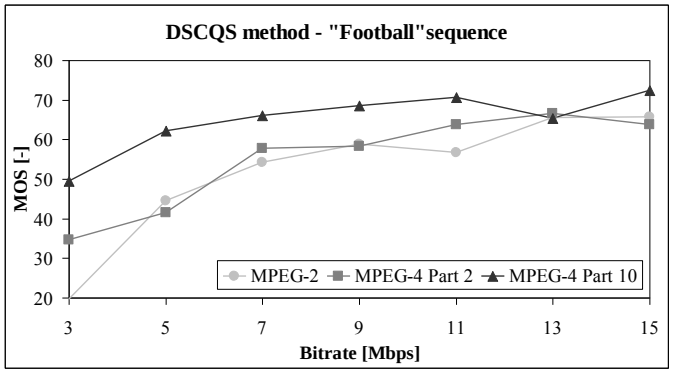


Figure 8. The results of the "Football" sequence using DSCQS method

By the ACR method 19 assessors were used (10 men and 9 women). Their age was in range from 13 to 71 years, the mean age was 31 years.

The figure 9 shows the results of the performed tests of the "Train" sequence using ACR method and the figure 10 shows the results of the performed tests of the "Football" sequence using ACR method.

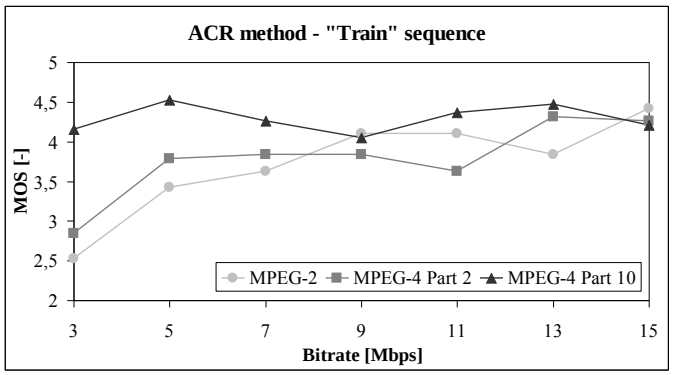


Figure 9. The results of the "Train" sequence using ACR method

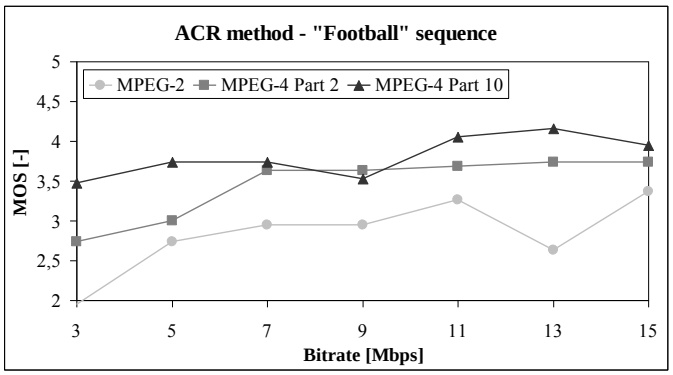


Figure 10. The results of the "Football" sequence using ACR method

V. CONCLUSION

In this article the subjective evaluation of the well-known MPEG compression standards (MPEG-2, MPEG-4 Part 2 and MPEG-4 Part 10) using DSIS, DSCQS and ACR methods was done. As could be supposed, according to the graphs, in all sequences H.264/MPEG-4 Part 10 compression standard can be considered as the best one. In the "Train" sequence with slow motion is the difference between the H.264 standard and the rest ones bigger than in the dynamic "Football" sequence. It can be due to the fact that the assessors perceived the difference of the video quality between the compression standards better in the scene with slow motion. The graphs also show that the assessors did not rate the video quality with extreme values. This testing is the first part of the research of a global video quality factor. This factor will rate not only the impact of a compression on the video quality but also other factors that influence the video quality, for instance the transmission link imperfection.

REFERENCES

[1] http://www.mpeg.org/.

[2] S. Winkler, Digital Video Quality: Vision Models and Metrics, John Wiley and Sons Ltd., 2005, 175 pages, ISBN 0-470-02404-6.

[3] J.-N. Hwang, Multimedia Networking: From Theory to Practice, Cambridge university Press 2009, 548 pages, ISBN 978-0-511-53364-8.

[4] J. Watkinson, The MPEG Handbook (second edition), Focal Press, 2004, 435 pages, ISBN 0-240-80578-X.

[5] C. Wooton, A Practical Guide to Video and Audio Compression, Elsevier Inc., 2005, 787 pages, ISBN 0-240-80630-1.

[6] E.G. Richardson, H.264 and MPEG-4 Video Compression, John Wiley and Sons Ltd., 2003, 281 pages, ISBN 0-470-84837-5.

[7] E.G. Richardson, The H.264 Advanced Video Compression Standard (second edition), John Wiley and Sons Ltd., 2003, 316 pages, ISBN 978-0-470-51692-8.

[8] Scientific Atlanta, "MPEG-4 part 10 AVC (H.264) video encoding," Scientific-Atlanta, June 2005, 19 pages, Part Number 7007887 Rev B.

[9] H.R. Wu, K.R Rao, Digital Video Image Quality and Perceptual Coding, Taylor and Francis Group LLC, 2006, 594 pages. ISBN 0- 8247-2777-0.

[10] RECOMMENDATION ITU-R BT.500-11, "Methodology for the subjective assessment of the quality of television pictures".

[11] http://www.irisa.fr/armor/lesmembres/Mohamed/Thesis/node2.html

[12] RECOMMENDATION ITU-T P.910, "Subjective video quality assessment methods for multimedia applications," 04/2008.

[13] http://media.xiph.org/vqeg/TestSeqences/ThumbNails/.

GA and PSO Based Feature Selection for Speaker Recognition Task

Michal Chmulik, Martin Hric, Roman Jarina
Dept. of Telecommunications and Multimedia
University of Žilina
Univerzitná 1, 01026 Žilina, Slovakia
michal.chmulik@fel.uniza.sk

***Abstract*—Besides the Genetic Algorithm, as a typical representative of evolutionary optimization techniques successfully used in many areas, relatively new approach to optimization called Particle Swarm Optimization (PSO) is presented as well as its variations Binary PSO and Angle Modulated PSO. We will mention basic principles of these algorithms with their advantages and disadvantages. In the experimental part we employed these 3 techniques for feature space optimization for the purposes of speaker recognition. Test results show that these algorithms are comparatively good in the meaning of accuracy and optimization process duration.**

Keywords- optimization techniques, genetic algorithms, particle swarm optimization, speaker recognition

I. INTRODUCTION

The optimization algorithms are very useful tools for solving many problems in various areas. Their significance increases rapidly in the cases, when the analytic solution is very complicated or even unreal. As an example we can present our case from experimental part. Feature space contains 69 MFC coefficients. Total number of features combinations is 269, so evaluation of each single combination is impossible even on the fastest available computer. Another example can be the traveling salesman problem. Optimization methods origin reaches to the fifties of the 20th century, when the basic principles were described by Turing et al. Another progress came in the seventies with the arrival of some new approaches to the optimization. The big explosion of optimization methods has achieved in the nineties and its popularity remains to the present. Generally, we can distinguish two basic categories of optimization algorithms [1-4]:

- deterministic
- stochastic

Deterministic methods, based on analytic mathematics methods, are useful only for simple functions that are typically unimodal. Deterministic techniques (simplex, gradient, Newton's method…) suffer from the fact that it mostly converges to the nearest local optimum and can not find the global optimum. Stochastic methods handle the convergence problem by exploiting the process of stochasticity. Solutions created by given optimization method are affected more or less by random process. Stochastic algorithms remain its stochasticity through the whole run of optimization process. Success and popularity of these algorithms dwell in the fact that it is possible to solve any problem that can be described by the criterion function. Contrary to the deterministic methods, this function can be much more complicated with many variables, nonlinear relations and with more global optima. Finding arguments of this function will lead us to the solution of the problem. With the criterion function is linked the term fitness that tells if the present problem solution is better than the previous. The higher fitness value detects better solution. The choice and the creation of the criterion function is the most important part without regard to used algorithm. In the present, there are many optimization techniques mostly inspired by animate or inanimate nature (genetic algorithms, evolution strategy, simulated annealing, differential evolution…). It is necessary to emphasize that there exists no one best method. The choice of algorithm always depends on given problem and method that can perfectly solve one problem, might be totally wrong and useless for any other problem. This fact is called NFT (No Free Lunch) theorem [1, 2].

II. GENETIC ALGORITHMS

Genetic algorithms (GA) [1-4] belong to the well-known methods and are one of the most frequently used stochastic optimization technique. The author is Holland who introduced this method in 1975 for the first time and nowadays exist more versions of GA. Its principle relates to the analogy with the nature and is based on the evolution laws described by Darwin and Mendel. The terminology is also taken from the field of biology. GA is built on cyclic pseudo-random generation of new solutions and natural selection. Individual represents one solution of given problem and usually is expressed as a bit string. Group of individuals creates a population. Genome is called a part of individual (bit or group of bits) with a certain meaning or function. Specific gene value is called allele. Creation of new individuals consists of two key elements:

- crossover
- mutation

Crossover describes the process when new offspring originate by combining the parent's genes. Crossover may happen at one or more points and the selection of this point is random. Mutation describes the process when randomly chosen

allele (one or more) is changed. Both processes – crossover, mutation – occur with a defined probability from the range of (0, 1). Mutation was designed to prevent the algorithm from premature convergence in local extreme. Another key part is choice of the parents. According to Darwin's laws, only best individuals survive, but worse individual can also contribute to future progress with proper genes. Selection is not pure random, fitness of individuals is taken into consideration. Common approach is a roulette-wheel selection. A fitness counts for every individual first and then the selection probability of given individual is directly proportional to its fitness value. Selection is usually simulated by roulette-wheel where the field size corresponds to the fitness of given individual. Selection probability of individual *i* is given by following equation:

$$T_i = \frac{G_i}{\sum_{i=1}^{n} G_i} \tag{1}$$

where *n* denotes number of individuals in the population and G_i is fitness of given individual. Other selection techniques are tournament or rank selection. Another improvement is the elitism that can accelerate algorithm run. In every population is chosen the best individual who is passed directly to new population and after that some kind of selection is occurred. More variations of GA have been created (messy GA, parallel GA, hybrid GA…). GA are useful for complex problems without knowledge about given problem. Next advantage is that bad solutions do not affect the final solution. The main disadvantage is that GA can converge prematurely to local optimum.

III. PARTICLE SWARM OPTIMIZATION

The Particle Swarm Optimization (PSO) was designed by James Kennedy and Russell Eberhart in 1995 [1, 2, 5]. This approach is inspired by social behavior of bird communities. PSO works as follows. Every individual from the community represents one single solution in the search area. Each of these individuals has its location in the search area, velocity directing its "flight" and remembers its own best location. The system is initialized by the population of random solutions (called particles) and tries to find the optimal solution creating new generations (populations) with better solutions. The fitness is counted for each randomly located particle. The particle with the best fitness stores its location into global "memory" (gBest). Except this memory shared by all particles, a personal memory of each particle also exists, in which the best individual solution is stored (pBest). Every particle in the population compares the present fitness value with its personal best and global best value, and if appropriate, replaces the value in the memory (personal or global) by the best one. When every particle knows gBest and pBest, the new velocity and the position is counted according to the following quotations:

$$v_d(t+1) = w\,v_d(t) + c_1\,rand\,(pBest_{id} - x_{id}(t)) + \\ + c_2\,rand\,(gBest_d - x_{id}(t)) \tag{2}$$

$$x_{id}(t+1) = x_{id}(t) + v_d(t+1) \tag{3}$$

where v_d *(t+1)* means the particle velocity in the next iteration (migration), v_d *(t)* is the particle velocity in the present iteration, x_{id} *(t+1)* means the particle position in the next iteration, x_{id} *(t)* is the particle position in the present iteration, $pBest_{id}$ denotes the best personal particle position found in the previous iterations, $gBest_d$ denotes the best global particle position in the population found in the previous iterations, w means the inertia weight, c_1, c_2 are the learning factors and rand means the random number from the interval (0, 1). Particles move to the new locations with the new velocities. New fitness of each particle is counted again and the whole cycle is repeated. Parameter w may be fixed or may vary linearly according the following formula:

$$w = w_{start} - \frac{(w_{start} - w_{end})\,it}{mig} \tag{4}$$

where *it* means the number of present iteration and *mig* denotes the total number of iteration. C_1 and c_2 influence the particle movement tendency (to *gBest*, *pBest* or to continue in the previous direction) only partially because these values are multiplied by random number. In some studies are products of random number and c_1 and c_2 substitute by fixed numbers hence the search stochasticity is removed. *PSO* is computational inexpensive and is very easy to implement. It also achieves a fast convergence, but on the other hand may tend to the premature convergence on a complex problem with many optima.

A. Discrete Binary PSO

Since many optimization problems are set in discrete space, Kennedy and Eberhart adapted original approach to the discrete binary version to allow working in this space [2, 6]. The difference between continuous and discrete *PSO* is in interpretation of the *velocity*. While in the original algorithm *velocity* determinates "real" change of position of particular particle in search space, in *binary PSO velocity* determinates the change probability of particular bit from the bit string creating given particle. Since the velocity in the meaning of probability must be from range (0, 1), sigmoidal logistic function is used to secure this. Higher values of vid will cause, that particle will tend to choose more likely 1 while with low values will be probably preferred 0. The discrete *binary PSO* can be described by the following equations:

$$x_{id}(t+1) = \begin{cases} 1, & \rho_{id} < s(v_{id}(t+1)) \\ 0, & else \end{cases} \tag{5}$$

$$s(v_{id}) = \frac{1}{1+\exp(-v_{id})} \quad (6)$$

where *i* denotes given particle, *d* denotes dimension, ρ_{id} is a vector of random numbers with uniform distribution and $s(v_{id})$ means sigmoidal function - value of v_{id} is counted according to the (2). Shortly, algorithm works as follows. For every individual – particle counts new velocity value v_{id} over each dimension. These values are based on the comparison of present values x_{id} with best values $gBest_d$ and $pBest_{id}$, similarly like in original *PSO*. Velocity values are substituted then into the (6) and the output from the sigmoidal function serves as a probability threshold to set the bit of each dimension to 1 or 0. This whole process repeats cyclically over the given number of generations.

B. Angle Modulated PSO

This version of PSO was designed by Pampura et al. in 2005 [7]. It is an interesting approach how to use standard *PSO* algorithm, working in continuous searching space, for solving optimization problems in discrete space. The authors took their inspiration in the field of signal processing from the telecommunications. In this approach, authors used angle modulation as a generating function that allows mapping continuous space to discrete. Generating function can be described as follows:

$$g(x) = \sin(2\pi(x-a)b\cos(A)) + d \quad (7)$$

$$A = 2\pi c(x-a) \quad (8)$$

where *x* is a single element from a set of uniform separated intervals determined by the desired bit string length, *a* and *d* represent horizontal and vertical shift respectively, *b* and *c* determinates maximal frequency of sin and cos function respectively. The standard *PSO* optimizes these variables *a*, *b*, *c*, *d* and then they are substituted back into the *angle modulation* function g(x). The resulting bit string creation is very easy – when output of the function *g(x)* has positive value, then corresponding bit is going to be 1, else 0.

IV. EXPERIMENTS

As a classification problem, the speaker recognition was chosen. The database, excerpted from MobilDat Sk [9,10], consisted from the set of 1100 speakers where every speaker had 3 samples. Every sample (recording) was uncompressed PCM wave file with sample frequency of 8 kHz with 16 bits depth and had length of 8 seconds. The database was parameterized by MFCC features - 23 MFCC + 23 delta MFCC + 23 delta delta MFCC. The window length at the parameterization process was 30 ms with a 10 ms overlapping. Frames that did not contain a speech were removed. In the experiments were used 2 samples as a reference or training data and 1 sample as a test data. As a classifier, we have chosen the k Nearest Neighbor (kNN) algorithm. Although this method is very simple, it remains its success in the data mining and is comparable with other more complicated methods and belongs to top 10 data mining algorithms [8]. We have chosen 20 classes for the tests so the kNN classifier should distinguish 1 between 20 speakers. All speakers were randomly selected from the database for every test. As a confidence measure for test evaluation was used precision, recall and f measure. Precision can be described as follows:

$$P = \frac{N_{CD}}{N_{TD}} \quad (9)$$

where N_{CD} denotes the number of correctly detected speakers and N_{TD} denotes the total number of detected speakers. The precision value expresses how many detected speakers truly belong to the given class. Recall, defined by the equation:

$$R = \frac{N_{CD}}{N_{TC}} \quad (10)$$

where N_{TC} means total number of correctly speakers to detect, says how many speakers were detected. Finally, f measure, described as follows:

$$F = \frac{2PR}{P+R} \quad (11)$$

is the harmonic mean of precision and recall and gives the overall classification rating. The precision, recall and f measure values are from the range of (0, 1). All tests were performed in the Matlab program on the PC with Intel® Core™ i5-760 2.80 GHz processor and 8 GB RAM. Each optimization algorithm – *GA, binary PSO and AMPSO* – ran in 100 iterations and the optimization (finding the best features combination) repeated 100 times for each algorithm. 10 best solutions for each algorithm were then picked and tested – every test consisted from 100 runs to secure the statistical credibility of obtained classified data. As a criterion function, mean f measure in association with mean number of correctly classified frames was used. In addition to this, solutions with lower features number were preferred. Parameters of particular algorithms have set as follows. For the *kNN* algorithm, the k value has set at 7. For *GA* we have chosen 25 individuals, 2 points crossover with crossover probability p_c = 0.85, mutation probability p_m = 0.05. For both *PSO – binary* and *angle modulated* – we have set the same parameters – 25 particles, inertia factor w = 0.729844, learning factors $c_1 = c_2$ = 1.49618, maximal velocity v = 4. In the Tab. 1 can be seen the best, the worst and the mean value of classification accuracy for each algorithm as well as mean time expressing how much time took the optimization process. Fig. 1 depicts the mean values of classification accuracy where the fourth value (80.16 %) was obtained by using all features. The number of used features chosen by optimization algorithms varied between 35 and 46. Here are examples of the "mask" of chosen features by particular algorithms in the same order as it is in the Tab. 1 and Fig. 1.

1111111111111011101111000011000000111101001100111100111001000000111

1111111101111110110110001100000110010110000101011000010000000110010001

1111011111111111111110000000010000000100101100001000101011100101101001

The classification accuracy with applying these features was 84.45 %, 84.49 % and 84.42 %.

TABLE I. CLASSIFICATION ACCURACY AND TIME

algorithm	worst [%]	best [%]	mean [%]	time [min]
AMPSO	81.38	84.49	83.15	43
bin PSO	83.41	85.09	83.97	47
GA	82.85	85.07	84.37	45

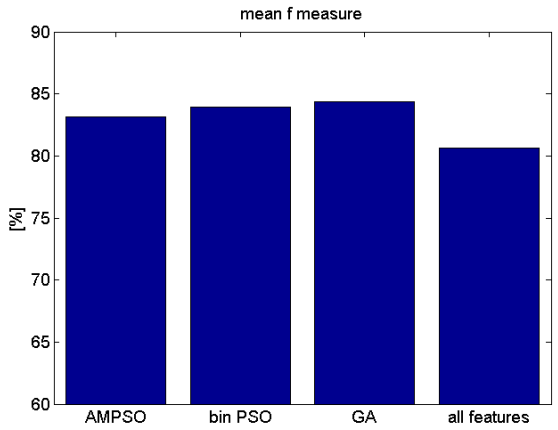


Figure 1. Classification accuracy

V. CONCLUSION

The results show that the best combination of features has been found by binary PSO, the worst by AMPSO. Considering the mean accuracy value, the best results achieved Genetic Algorithm. AMPSO scored slightly worse classification accuracy than the other 2 algorithms. Talking about the time consumption, the fastest algorithm is AMPSO, but all 3 algorithms are comparable. As it can be seen on Fig. 1, all algorithms have found features combination, that gives better classification accuracy in comparison with the case when all features were used. The results also show that there exist more "optimal" features combinations. In the future work, we will focus on MPEG-7 audio features and will employ these 3 optimization algorithms to find "optimal" combination of features for speaker recognition as well as for general sound recognition.

ACKNOWLEDGMENT

The work is supported by the Slovak Research and Development agency under the contract No. APVV-0369-07.

REFERENCES

[1] ZELINKA I., OPLATKOVÁ Z., ŠEDA M., OŠMERA P., VČELAŘ F.: Evoluční výpočetní techniky - principy a aplikace, BEN - technická literatura, 2008

[2] KENNEDY J., EBERHART R. C.: Swarm intelligence, Academic Press, 2001

[3] BÄCK T., FOGEL D. B., MICHALEWICZ Z.: The handbook of evolutionary computation, IOP Publishing Ltd and Oxford University Press, 1997

[4] ASHLOCK D.: Evolutionary computation for modeling and optimization, Springer, 2005

[5] KENNEDY J., EBERHART R.: Particle swarm optimization, Proceedings of IEEE International Conference on Neural Networks, 1995, vol.4, no., pp.1942-1948

[6] KENNEDY J., EBERHART R.: A discrete binary version of the particle swarm optimization algorithm, IEEE International Conference on Systems, Man, and Cybernetics, 1997, vol.5, no., pp.4104-4108

[7] PAMPARA, G.; FRANKEN, N.; ENGELBRECHT, A.P.: Combining particle swarm optimisation with angle modulation to solve binary problems, IEEE Congress on Evolutionary Computation, 2005, vol.1, no., pp.89-96

[8] W. XINDONG, K. VIPIN: The Top 10 Algorithms in Data Mining, Chapman & Hall/CRC, 2009

[9] RUSKO M, TRNKA M., DARJAA S.: MobilDat-SK – a Mobile Telephone Extension to the SpeechDat-E SK Telephone Speech Database in Slovak, Proc. of SPECOM, 2006, St. Petersburg

[10] JUHÁR J., ONDÁŠ S., ČIŽMÁR A., JARINA R., RUSKO M., ROZINAJ G..: Development of Slovak GALAXY / VoiceXML Based Spoken Language Dialogue System to Retrieve Information from the Internet, Proc. Interspeech 2006, Pittsburg, USA, 2006, pp. 485-488.

Fourier and Wavelet Analysis – an engineers approach

Jan Skapa, Marek Penhaker, Jan Latal, Petr Koudelka, Jan Vitasek
Department of Telecommunications, Faculty of Electrical Engineering and Computer Science, VŠB—TU Ostrava, 17. listopadu 15, 708 33, Ostrava—Poruba
jan.skapa@vsb.cz

***Abstract* — This article discusses advantages and disadvantages of using Fourier and wavelet transform for time--frequency or shift—scale analysis of one—dimensional signal. The use of wavelet transform is discussed in very well known terms of Fourier analysis to emphasize the advantages of wavelet transform according to Fourier transform.**

***Index Terms* — DSP, Fourier transform, wavelet transform.**

I. Introduction

Suppose that we have a record of piano music on a CD. If we want to find out, which tones and chords are played through the composition (and we do not have absolute pitch), we could use the apparatus of DSP. The Fourier transform perfectly fits to this purpose, saying which frequencies are included in the record. If we know, that relation between two neighboring chromatic tones is $\sqrt[12]{2}$, there is very straight way to determine, which tones were played during the record. The problem with Fourier analysis is that it does not say, when the tone (frequency) starts and ends to sound. It only says, that the sound has been played. To determine when the individual tones started to sound sometimes spectrograms are used. But using spectrograms the spectral resolution is reduced. This problem is resolved by use of wavelet transform, which can provide time-frequency-like analysis in a bit different way.

II. Used notation

In this work we suppose using the DSP techniques, so we are working with digitalized signals. The sampling frequency is denoted f_s, the related sampling period is denoted ΔT. The signal is denoted $f[n\Delta T], n\in\mathbb{Z}$, which means that the samples were taken in discrete moments with distance ΔT between individual samples. Usually the shorter form is used denoting the signal as $f[n]$. We denote the length of signal N (number of samples contained in the signal). The frequency resolution is defined $\Delta f = f_s / N$. It means that we can not distinguish frequencies, which are closer than the frequency resolution is.

III. Discrete Fourier transform

Discrete Fourier transform map data from the time (or spatial for example) domain (the original or input data vector) to the frequency domain. The Fourier analysis in discrete form (DFT) is defined

$$f[n\Delta T]=\sum_{k=0}^{N-1}\hat{f}[n\Delta f]\mathrm{e}^{2\pi\frac{nk}{N}},$$

$$\hat{f}[n\Delta f]=\Delta T\sum_{k=0}^{N-1}f[n\Delta T]\mathrm{e}^{-2\pi\frac{nk}{N}}.$$

IV. Spectrogram

Spectrograms are usually created in that way that we take a short part of the signal with length of $2^k, k\in\mathbb{N}$ points. Then we multiply the taken signal with a window, which is used to minimize the Gibbs phenomenon. The Fourier transform is computed on specified number of points, then the window is shifted about specified number of points and the same process starts again. Spectra arising from the Fourier transform are placed side-by-side to create time—frequency image of the signal. The problem arising from this approach is the spectral resolution reduction. If we take short parts of the signal, which is created with specified sampling frequency, the windowed signal has frequency resolution $\Delta f_w = f_s / N_w$, where N_w is the length of the window (the length of the short part of signal).

V. Discrete wavelet transform

Discrete wavelet transforms map data from the time domain (the original or input data vector) to the wavelet domain. The DWT of a signal *x* is calculated by passing it through a series of filters. First the samples are passed through a low pass filter with impulse response $h[n]$ resulting in a convolution of the two:

$$cA_1[n]=(f*h)[n]=\sum_{k=-\infty}^{\infty}f[k]h[n-k].$$

The signal is also decomposed simultaneously using a high-pass filter $g[n]$. The outputs giving the detail coefficients (from the high-pass filter) and approximation coefficients (from the low-pass). It is important that the two filters are related to each other and they are known as a quadrature mirror filter, which means

$$|H(e^{j\Omega})|^2+|G(e^{j\Omega})|^2=1.$$

However, since half the frequencies of the signal have now been removed, half the samples can be discarded according to Nyquist's rule. The filter outputs are then subsampled by 2:

$$cA_1[n]=(f*h)[n]=\sum_{k=-\infty}^{\infty}f[k]h[2\mathrm{n}-k],$$

$$cD_1[n]=(f*g)[n]=\sum_{k=-\infty}^{\infty} f[k]g[2n+1-k] \quad .$$

With the subsampling operator $\downarrow$, defined by

$$(x\downarrow k)[n]=x[kn] \quad ,$$

the above summation can be written more concisely.

$$cA_1=(f*h)\downarrow 2 \quad ,$$

$$cD_1=(f*g)\downarrow 2 \quad .$$

There are two properties, that hold for the filter coefficients, $\sum_{\forall k} h^2[k]=1$. That means, that the scaling function has energy 1. Another property is $\sum_{\forall k} h[k]=\sqrt{2}$. For the wavelets (high-pass filters) holds $\sum_{\forall k} g^2[k]=1$, $\sum_{\forall k} g[k]=0$.

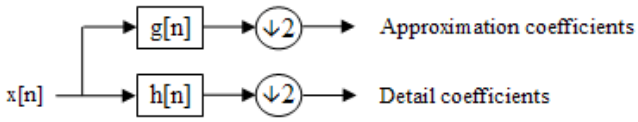


Fig. 1. Dyadic wavelet decomposition.

The wavelet decomposition is then formed by the approximation and detail coefficients,

$$f\rightarrow(cA_1, cD_1) \quad .$$

This step, filtration and downsampling divides the spectral parts of the signal into two parts, the approximation part contains the lower frequencies, the detailed part contains the higher frequencies. The important note is, that the spectra of this parts overlaps.

A. Cascade decomposition

The decomposition is repeated to further increase the frequency resolution and the approximation coefficients decomposed with high and low pass filters and then down-sampled. This decomposition is called multiresolutional analysis.

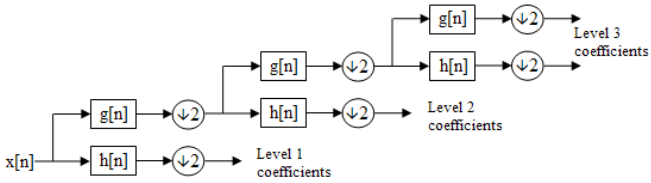


Fig. 2. Multiresolutional analysis.

This is represented as a binary tree with nodes representing a sub-space with a different time-frequency localization.

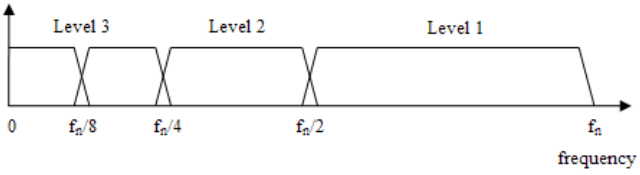


Fig. 3. Spectrum divided by wavelet filters.

B. Conservation of energy

One of the important properties of the discrete wavelet transform is that it conserves the energies of the signal. The energy of the signal is defined by

$$E=\sum_{k=0}^{N-1} f^2[k] \quad .$$

The energy of the signal should be equal to the energy of the wavelet decomposition,

$$E=\sum_{k=0}^{N/2-1} cA^2[k]+\sum_{k=0}^{N/2-1} cD^2[k] \quad .$$

If the multiresolutional analysis is used, the conservation of energy still holds.

C. Compaction of energy

The energy of the trend (approximation) coefficients cA accounts a large percentage of the energy of the transformed signal (cA, cD). Compaction of energy will occur whenever the magnitudes of the fluctuation (detail) values are significantly smaller then the trend (approximation) values. According to this property the small fluctuations feature of the wavelet transform should be recalled. If the samples of the signal f are samples of a continuous analog signal with a very short time increment between the samples, then the neighboring samples will be close to each in value and the differences between them will be close to 0. The detail coefficients will be close to 0 too and almost all the energy of the signal will be contained in the approximation part. This feature is used in signal compression using wavelet transform. If we suppress those coefficients, whose values are close to 0, we can store or transmit only nonzero coefficients. If the small fluctuations feature holds (the detail coefficients are close to zero in value), we can transmit only the approximation part, which has $N/2$ samples in length, so it provides 50 % compression.

D. Mallat's cascade algorithm in matrix form

The cascade algorithm in matrix form can be written as follows

$$\begin{bmatrix} cA \\ cD \end{bmatrix}=\left[\begin{bmatrix}\begin{bmatrix} h_3 \\ g_3 \end{bmatrix} h_2 \\ g_2 \end{bmatrix} h_1 \right][x] \quad ,$$

where h and g are transormation matrices for each level of decomposition. We can determine each decomposition filter by multiplying inner matrices.

VI. Experiments

For experiments with wavelet decomposition I have used the signal defined as

$$f(t)=20t^2(1-t)^4\cos(12\pi t) \quad .$$

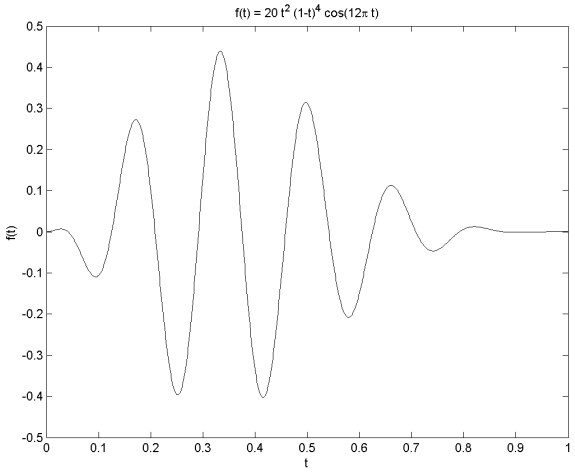


Fig. 4. Signal $f(t)=20t^2(1-t)^4\cos(12\pi t)$.

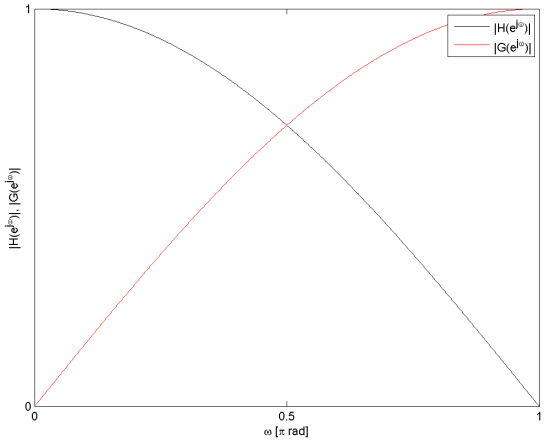


Fig. 5. The magnitude characteristics of the wavelet filter db1.

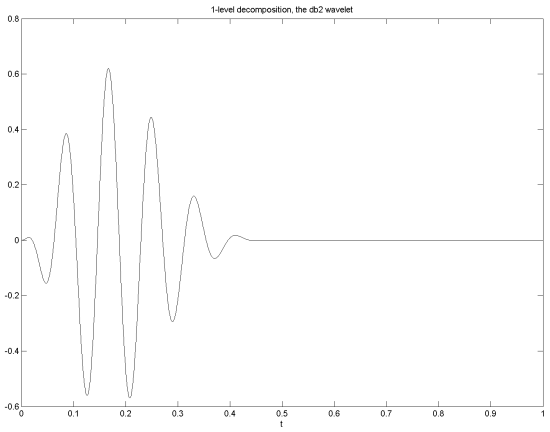


Fig. 6. 1-level decomposition, wavelet db2.

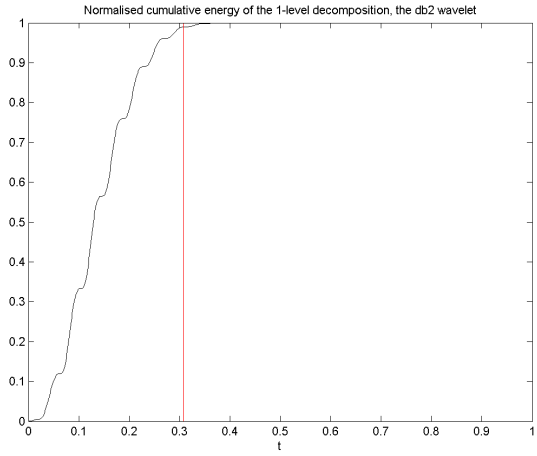


Fig. 7. Normalized cumulative energy profile of the 1-level decomposition, wavelet db2.

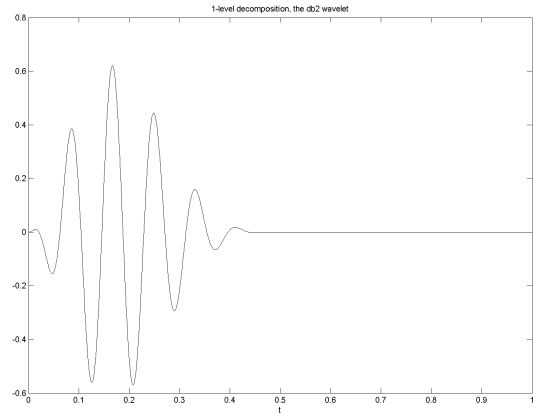


Fig. 8. 1-level decomposition, wavelet db2.

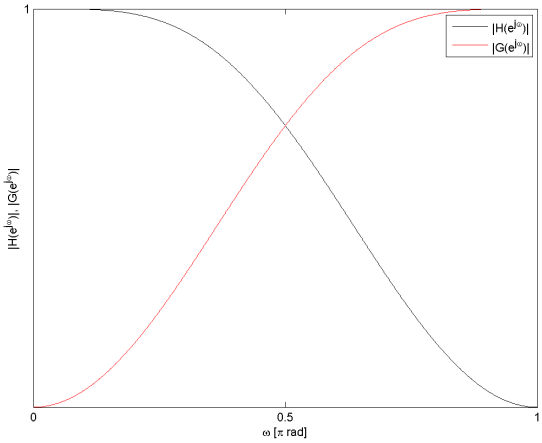


Fig. 9. The magnitude characteristics of the wavelet filter db2.

If we use wavelet transform to compress a 2-dimensional signal (an image), we get very good results comparing with compression based on Fourier transform. In JPEG compression is used DCT – Discrete Cosine Transform, which is a special case of the Discrete Fourier Transform for evenly periodized signals.

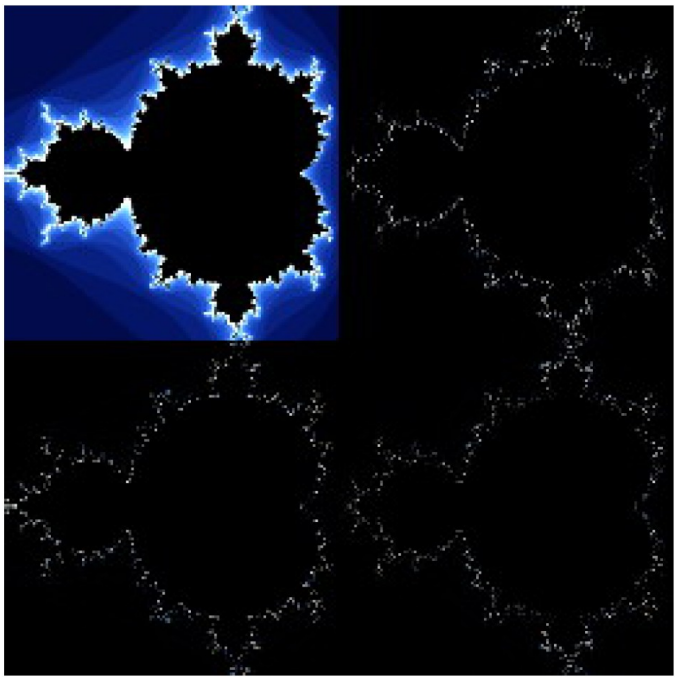


Fig. 10 Decomposition with “db2” wavelet

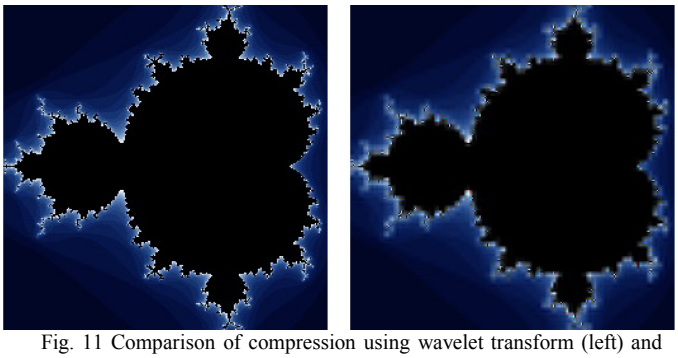


Fig. 11 Comparison of compression using wavelet transform (left) and JPEG – Discrete Cosine transform (right). Both pictures have the same size in Bytes.

VII. Conclusion

The article shows how wavelets can be used as a tool for signal compression and denoising. The filtering approach is used to explain wavelet transform to those people, who are familiar with commonly used Fourier transform and convolution filtering.

VIII. Acknowledgement

This project has been supported by the Czech science foundation grant GA 102 09/550.

References

[1] Vidakovic B. Statistical modeling by wavelets. WSPS, Wiley, 1999. ISBN 0471293652, 389p.

[2] Bachman, G., Narici, L., Beckenstein, E. *Fourier and wavelet analysis*. New York: Springer, 2000. ISBN 0-387-98899-8.

[3] Briggs, W. L., Henson, V. E. *The DFT: an owner's manual for the discrete Fourier transform.* Philadelphia: Society for Industrial and Applied Mathematics, 1995. ISBN 0-89871-342-0.

[4] Daubechies, I. *Orthonormal bases of compactly supported wavelets.* Wiley, 1988.

[5] Daubechies, I. *Ten lectures on wavelets.* Philadelphia: Society for Industrial and Applied Mathematics, 1992. ISBN 0-89871-274-2.

[6] Frazier, M. W. *An introduction to wavelets through linear algebra.* New York: Springer, 1999. ISBN 0-387-98639-1.

[7] Chui, CH. K. *Wavelets; A Mathematical Tool for Signal Processing.* Philadelphia: SIAM, 1997. ISBN -0-89871-384-6.

[8] Strang, G., Nguyen, T. *Wavelets and filter banks.* Wellesley: Wellesley--Cambridge Press, 1996. ISBN 0-9614088-7-1.

Optical Communications

chairman:

Leos Bohac

The Effective Bandwidth Utilization in the Hybrid PON Networks

Rastislav Roka
Institute of Telecommunications
Slovak University of Technology
Bratislava, Slovakia
rastislav.roka@stuba.sk

Abstract - **In this contribution, we introduce principles of the hybrid PON networks. First, we focus on algorithms of the dynamic bandwidth allocation that are used in the TDM-PON networks. Also, basic features that must be considered at the simulation of these algorithms are specified. Then, we present algorithms of the dynamic wavelength allocation that are used in the WDM-PON networks. Based on possible evolution trends of the HPON network, we can suppose an extending interest about bandwidth- and/or wavelength- management protocols. Therefore, studies for the effective bandwidth utilization in the hybrid PON networks will be continued.**

Keywords **- the hybrid PON network, DBA algorithms, DWA algorithms.**

I. INTRODUCTION

In this contribution, we focus on the effective bandwidth utilization in the hybrid PON networks - smooth transition between TDM-PON and WDM-PON networks. A main reason for this motivation is that at first a number of operating TDM-PON networks is even now high and is still rising and at second a utilization of installed optical infrastructures is maximizable for transmission capacity's increasing. On the other side, a creating the HPON network can be used as the upgrade of old networks in many cases with a utilization of relevant parts of the original infrastructure with minimum financial costs. It's a question whether a consistent transition from the TDM-PON to the HPON network only is requested or further to the WDM-PON network is suitable. Therefore, it is important to identify maximum effective bandwidth utilization in the HPON network.

First, we introduce principles of the hybrid PON networks. Then, algorithms of the dynamic bandwidth allocation commonly used in the TDM-PON networks are focused. Also, some basic features that must be considered at the simulation of these algorithms are simply specified. Finally, algorithms of the dynamic wavelength allocation used in the WDM-PON networks are presented. Based on possible evolution trends of the HPON network, we can suppose an extending interest about bandwidth- and/or wavelength- management protocols. Therefore, studies of new dynamic wavelength and bandwidth allocation algorithms for the effective bandwidth utilization in the hybrid PON networks will be needed to be continued.

II. PRINCIPLES OF THE HYBRID PASSIVE OPTICAL NETWORK

The HPON presents a hybrid network as a necessary phase of the future transition from TDM to WDM passive optical networks. The HPON network utilizes similar or soft revised topologies as TDM-PON architectures. For the downstream and upstream transmission, TDM and WDM approaches are properly combined, i.e. it is possible to utilize the time-division or wavelength-division multiplexing of transmission channels in the common passive optical network.

In present days, PON networks with the TDM multiplexing technique are creating in many countries, including also Slovakia. In the near future, we can expect different motivations for developing of HPON networks:

1. Creating the new PON network overcoming TDM-PON network possibilities with minimum financial costs. In this case, various optical resources are utilized, but it isn't the full-value WDM-PON network. Also, the TDM approach is still utilized from various reasons. So created HPON network is not very expensive (in a comparison with the WDM-PON network) and provides a sufficient transmission capacity for customer needs in a long-time horizon.

2. Preparing the transition from the TDM-PON to the WDM-PON network with minimum costs for rebuilding of the existing TDM-PON infrastructure. Such HPON network should satisfy following features:

- a backward compatibility with the original TDM-PON architecture, a coexistence of TDM and WDM approaches,
- a maximum exploitation of the existing optical infrastructure, optical fibers and optical equipments,
- new bonus functions for the network protection and the fast traffic restoration in a case of failures.

Our HPON network model [1] is based on principles of the evolutionary architecture from the TDM-PON network utilizing few WDM components (Fig.1). A basic architecture consists of the one-fiber topology and several topological links connected to the optical network terminals (ONT) through the remote nodes (RN). Logically, a connection of the point-to-point type is created between the optical line terminal (OLT) and the RN. The RN consists of either a passive optical splitter (TDM-PON) or an arrayed waveguide grating (WDM-PON). As resources of optical radiations in the OLT, tunable lasers (TL) are utilized in order to decrease a number of necessary optical sources. A number of tunable lasers in the OLT is smaller than a number of transmission channels utilized in a network,

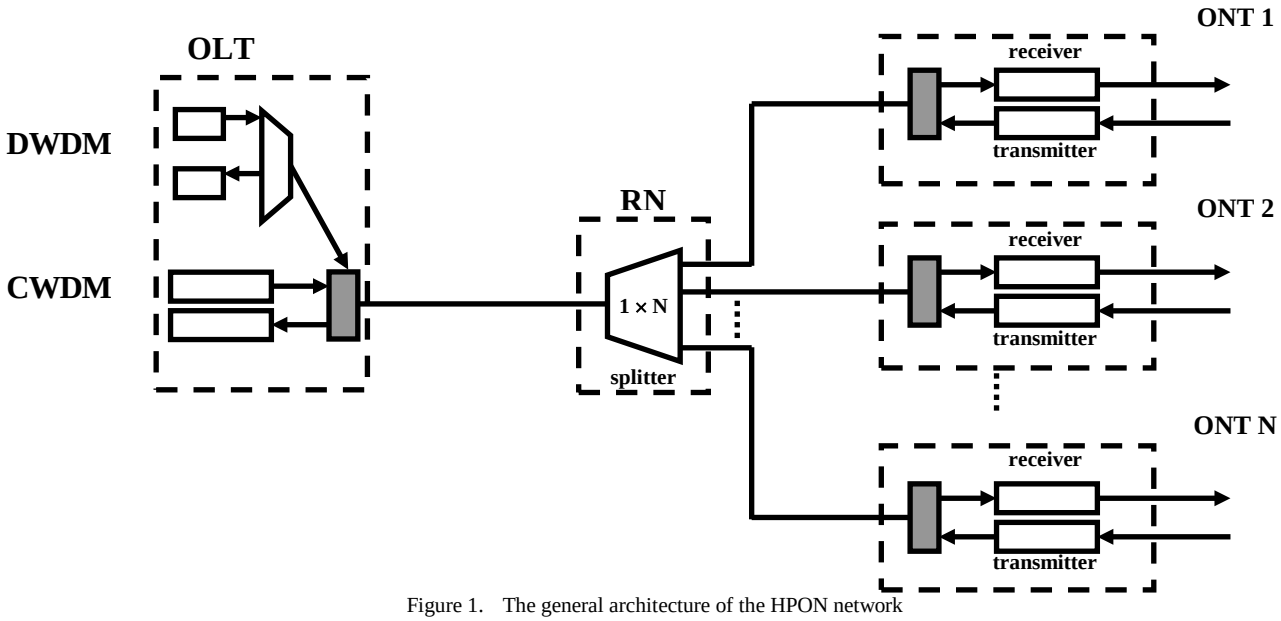


Figure 1. The general architecture of the HPON network

therefore various subscribers can dynamically share tunable lasers. Of course, there are also other possible HPON architectures [2], [3], [4] that can be later also included in this HPON network model.

At present days, two basic development trends of WDM systems are boosted. First approach is the dense WDM (DWDM). Conventional DWDM systems have a spectral spacing of 50 or 100 GHz (corresponding to 0,4 nm or 0,8 nm at wavelengths near 1,5 μm). Second approach is the coarse WDM (CWDM). CWDM systems promise all the key characteristics of network architectures (transparency, scalability and low cost), especially for building a metro/access network [5]. The metro/access network does not require the same bandwidth and distance requirements as the long-haul network. The challenge for metro/access networks is in distributing the capacity delivered by the long-haul network core and in aggregating from the network edge back to the long-haul core.

Next, we can present basic steps of a consistent transition from the existing TDM-PON network to the our HPON network model:

In the first step, a change in the OLT equipment is executed, specifically with adding a splitter to combine the original CWDM system and the additional DWDM system into one-fiber transmissions.

In the second step, appropriate changes in the ONU equipments are executed, specifically with adding CWDM or DWDM filters assigned to specific wavelengths for ONU transceivers.

In the CWDM OLT node, a practical hardware implementation of DWDM elements on the CWDM network requires a few changes [5]:

1) to add the DWDM multiplexer/demultiplexer between the splitter and optical receivers,
2) to change CWDM GBICs (Gigabit Interface Converter) with the number of DWDM GBICs corresponding to the size of the DWDM multiplexer that was added,
3) to make corresponding multiplexer and transceiver changes at the OLT node.

I. Algorithms of the DBA in the TDM-PON Network

In the PON with the TDM multiplexing technique, a static allocation of the bandwidth to individual subscribers is very inefficient. Therefore, for more efficient utilization of the available wavelength λ, it is necessary to consider about algorithms of the dynamic bandwidth allocation. The developing of the multi-point control protocol (MPCP) is focused to dynamically assign the upstream bandwidth in optical fibers, however does not specify any particular dynamic bandwidth allocation (DBA) algorithm. Instead, it is intended to facilitate the implementation of any DBA algorithm.

The DBA algorithms [6] can be used in the DBA module of the MPCP arbitration mechanism for calculation and designation of the collision-free upstream transmission schedule of ONU units and for generation of GATE messages accordingly. The DBA algorithms can be categorized into algorithms with statistical multiplexing and algorithms with quality of service (QoS) assurances (Fig. 2). In the first group, the interleaved polling with adaptive cycle time (IPACT) method and the control theoretic extension (CTE) of this IPACT method are included. The second group is further subdivided into algorithms with absolute QoS assurances – e.g. the bandwidth guaranteed polling (BGP) method and the deterministic effective bandwidth (DEB) method – and algorithms with relative QoS assurances – e.g. the DBA for multimedia approach and the DBA for QoS method.

The developed DBA schemes address two key challenges – accommodating traffic fluctuations and providing QoS to the traffic. The first challenge has been addressed by estimating the amount of traffic arriving between two successive transmission grants to the ONU using a proportional control in the IPACT enhancement and a one-step prediction in the DBA for QoS method.

For future, it is possible to develop and evaluate more sophisticated prediction mechanisms for the DBA and more complex control mechanisms.

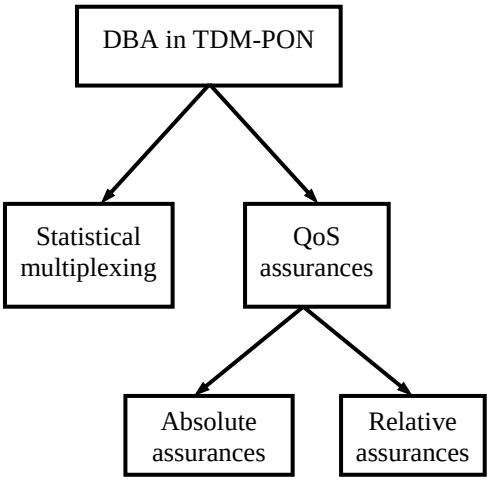


Figure 2. The taxonomy of dynamic bandwidth allocation algorithms

The second challenge has been addressed by a wide variety of QoS mechanisms that have generally been adapted form the QoS mechanisms developed for the Internet. These approaches fall into two main categories of absolute assurances and relative assurances. The absolute assurance mechanisms are designed to provide lossless service with deterministic delay bounds. On the other hand, the relative assurance mechanisms differentiate among QoS levels provided to the different traffic classes. For future, a consideration is focused on the development of QoS mechanisms that provide absolute statistical QoS assurances.

II. CONSIDERED SIMULATION FEATURES

The transmission scheme introduced on Fig. 3 presents a data traffic in the upstream direction with the service cycle time T_p, the transmission window for given ONUs T_i and the guard interval T_g. This transmission scheme is outgoing from principles of the MPCP mechanism where bandwidth units for particular ONUs are allocated by the OLT unit. First, the OLT can receive REPORT messages particularly from each ONU and consequently sent out the GATE message to the specific ONU with information about the time slot allocation. By this way, the collision-free upstream transmission was secured but inefficiency. Later, a method with the bandwidth allocation decision after receiving REPORT messages from all ONUs in one service cycle time was implemented. After calculation, the OLT send out GATE messages according to applied allocation method. At the method design, only one service cycle T_p is considered.

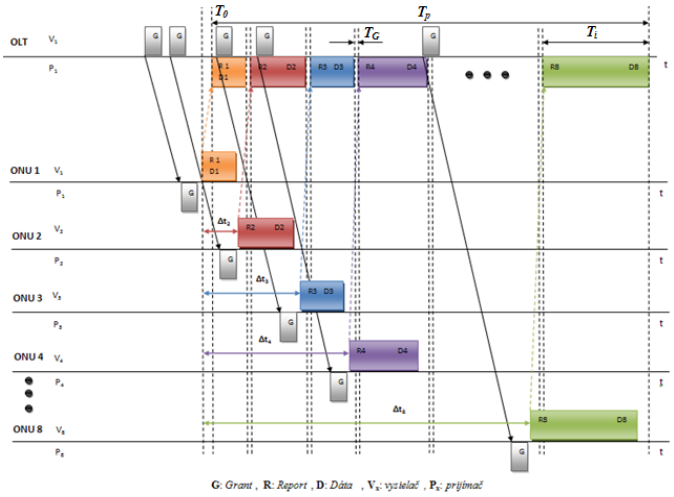


Figure 3. The transmission scheme of data traffic in the upstream direction

Following considerations must be specified in input parameters of the simulation model:

1) The total upstream bandwidth P_{total} is a theoretical value given by characteristics for optical networks. In praxis, the maximum upstream bandwidth for given ONUs is defined and given by SLA parameters. This limit can be for all ONUs equal or not. There can be ONUs with different requirements on the transmission bandwidth. There exist two possibilities – if the sum of all requests P_{ONU} exceeds the total upstream bandwidth P_{total} or the sum of all requests P_{ONU} is lower than the total upstream bandwidth P_{total}. From this result constraints on the maximum upstream P_{ONU}.
2) The maximum distance of ONUs from the OLT is defined in the ITU-T G.983 Recommendation.
3) The guard interval T_q is utilized for separating frames from different ONU transmissions. The T_q allows setting up a receiver for data receiving from following ONU and compensates RTT delays of particular ONUs.
4) Based on the ONU bandwidth request, the transmission window T_i can be simply designate for particular ONUs as following:

$$T_{i,ONU} = \frac{P_{i,ONU}}{P_{total}} \quad (1)$$

The OLT receiver is setting for data receiving in a defined time. The GATE message contains time stamp information for synchronizing internal local clocks of each ONU. This is on principle displayed on Fig. 3 where transmission times are assigned in consideration of the RTT.

5) The average frame delay Δt_{aver} can be expressed using a formula

$$\Delta t_{aver} = \frac{\sum_{i=1}^{N} \Delta t_i}{N} \quad (2)$$

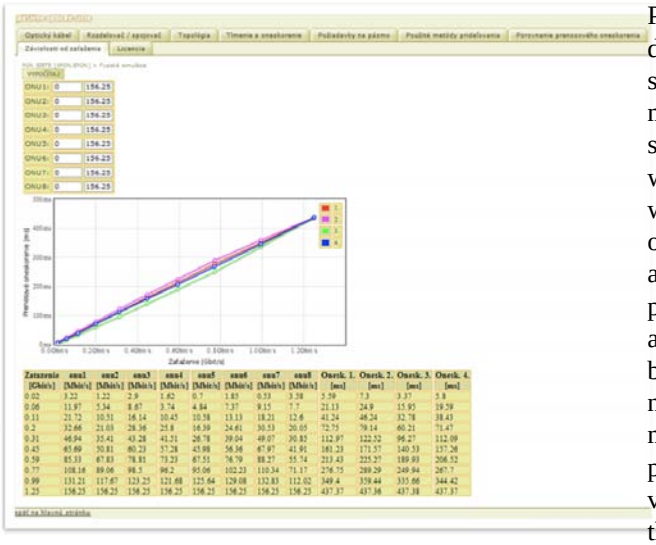


Figure 4. The transmission loading for various allocation methods

The bandwidth requests and transmission windows can be equipped with various methods. For comparing these allocation methods, a graphical representation of the transmission loading can be useful (Fig. 4). As we can see, changes of transmission delays can be observed under influence of changing bandwidth requests.

III. Algorithms of the DWA in the WDM-PON Network

In the PON with the WDM multiplexing technique, various wavelength or bandwidth management protocols are studied. The using of these protocols is depending on the dependency between used wavelengths [3]. If each channel works independent of the others, then the WDM-PON system may not need an additional medium-access control (MAC) protocol. In the initial stage of the WDM-PON implementation, this protocol-transparency property is regarded as one of advantages of the WDM-PON network. But this scheme may not utilize the available bandwidth efficiently, especially when some wavelengths are overloaded while others have a light load. However, if used wavelengths are shared among ONU units, then the total throughput may be increased significantly. In addition to the basic communication protocol, there are also needed new dynamic wavelength and bandwidth assignment algorithms for assigning not only time slots but also wavelengths to each OLT and ONU unit. These algorithms for the WDM-PON network are in their initial stage of study because they are closely dependent on the architecture, which is still in the early phase of study.

IV. Conclusion

In this contribution, we introduced principles and various aspects of the increasing of the bandwidth efficiency in the hybrid PON networks. At first, we focused on algorithms of the dynamic bandwidth allocation that are used in the TDM-PON networks. For future, a consideration is focused on the development of QoS mechanisms that provide absolute statistical QoS assurances. Also, some basic features that must be considered at the simulation of these algorithms are simply specified. Finally, algorithms of the dynamic wavelength allocation used in the WDM-PON networks were shortly presented. Based on possible evolution trends of the HPON network, we can suppose an extending interest about bandwidth- and/or wavelength- management protocols. Therefore, studies of new dynamic wavelength and bandwidth allocation algorithms for the effective bandwidth utilization in the hybrid PON networks will be needed to be continued. These studies must incorporate many possible network architectures that are still in the early phase of development. No specific one is dominant between various WDM-PON architectures. Therefore, in addition to the basic communication protocols, new dynamic wavelength and bandwidth assignment algorithms are needed.

Acknowledgment

This work is a part of research activities conducted at Slovak University of Technology Bratislava, Faculty of Electrical Engineering and Information Technology, Department of Telecommunications, within the scope of the project VEGA No. 1/0106/11 “Analysis and proposal for advanced optical access networks in the NGN converged infrastructure utilizing fixed transmission media for supporting multimedia services”.

References

[1] R. Róka: The designing of passive optical networks using the HPON network configurator, In: International Journal of Research and Reviews in Computer Science – IJRRCS, Vol. 1, No. 3, 2010, pp. 38-43, ISSN 2079-2557

[2] L. G. Kazovsky, W.-T. Shaw, D. Gutierrez, N. Cheng, S.-W. Wong: Next-generation optical access networks, Journal of Lightwave Technology, vol. 25, no. 11, November 2007, pp. 3428-3442

[3] A. Banerjee, Y. Park, F. Clarke, H. Song, S. Yang, G. Kramer, K. Kim, B. Mukherjee: WDM-PON technologies for broadband access – a review, Journal of Optical Networking, vol. 4, no. 11, November 2005, pp. 737-758

[4] F.-T. An, D. Gutierrez, K. S. Kim, J. W Lee, L. G. Kazovsky: SUCCESS-HPON: A next-generation optical access architecture for smooth migration from TDM-PON to WDM-PON, IEEE Comm. Magazine, vol. 43, no. 11, November 2005, pp. S40-S47

[5] R. Róka: The utilization of the DWDM/CWDM combination in the metro/access networks, In: SympoTIC 2003 – Joint 1st Workshop on Mobile Future & Symposium on Trends in Communications, Bratislava (Slovakia), 26.–28.10.2003, pp. 160-162, ISBN 0-7803-7993-4

[6] M. P. McGarry, M. Maier, M. Reisslein: Ethernet PONs – a survey of dynamic bandwidth allocation (DBA) algorithms, IEEE Communications Magazine, August 2004, pp. S8-S15

The Labelled Optical Burst Switching OMNeT++ Model With Accurate Time Offset Evaluation

Milos Kozak, Leos Bohac
Faculty of Electrical Engineering, Department of telecommunication technology
Czech Technical University in Prague
Prague, Czech Republic
milos.kozak@fel.cvut.cz, bohac@fel.cvut.cz

***Abstract*—in this paper, problems of labelled optical burst switching are discussed. In purely optical networks a lightpath along which optical bursts are consequently transferred must be first dynamically set up using optical switches in nodes and switching information contained in the burst header itself. However, creation of the lightpath takes time due to technological limits of current switching elements. This time must be considered in a design of signalling protocol otherwise the data packets located at the front of the optical burst and sometimes even further are going to be lost. The OMNeT++ simulation models are proposed to solve these problems. The simple mathematical model was created as well to evaluate correct time offset required between optical burst header and its data payload that follows it. Keeping in mind a correct offset in all optical nodes facilitates establishment of optical burst lightpath where no data packet losses are incurred. Furthermore a problem of inappropriate time offset and its impact on data loss with respect to a variety of modern phase optical modulations is discussed. This analysis yields the demand of ultra fast OXC switching matrix with switch time in scale of nanoseconds.**

Keywords-optical burst switching, graph theory, time offset, all-optical networks

I. INTRODUCTION

The massive deployment of the access technologies such as Asymmetric Digital Subscriber Line (ADSL), wireless networks or even passive optical networks in cities have facilitated a variety of new multimedia services such as YouTube, Skype, etc. But the deployment of these services led to a significant increase in the data traffic. As a result the transmission and switching systems must be modernised to handle this massive amount of user data transmitted to and from data centers where servers of these services are located.

A number of new technologies have recently emerged to tackle the aforementioned problems. The first one is the wavelength division multiplexing (WDM) which is based on usage of more channels in one fibre. The foreseen speed limitations of the electronic switching and routing systems have stimulated active research on promising technologies like optical packet switching (OPS) and optical burst switching (OBS). The main focus of this article is optical burst switching network modelling in simulation environment of OMNeT++. The OBS is an all-optical switching technology considered as a candidate for the next-generation transmission systems. The slow speed of recent all-optical switching elements and processing units prohibits full optical processing. As a consequence it is still necessary to evaluate the content of the switching header in real time and drive the optical switching elements as quickly as possible to avoid losing the data content of the burst that follows a header. It is not yet possible to delay an optical burst arbitrarily in a buffer for given period of time as we do not have an optical memory.

What seems to be feasible today is to use in-advance reservation mechanism. In this case the lightpath is set before an optical burst is sent and torn down if it is not needed any more. It means that no optical buffering is needed. An optical burst is usually defined as a number of continuous packets destined to a common egress point [1]. The burst size can vary from a single IP packet to a large data set at milliseconds time scale. The small time scale allows fine grained multiplexing over a single wavelength. Every optical burst is preceded by burst header packet (BHP) which is set onto a different, service, wavelength and is used for signalling. BHP carries information significant for switching node to set an appropriate optical cross connect (OXC) to an appropriate state to satisfy reservation and routing demands. Fundamental parts of switching node are depicted in Figure 1. From Figure 1. it is clear there are two different plains, control and data plane. The control plane is point-to-point and the optical-electrical-optical conversion is used because BCP processor is an ordinary electronic device conducting OXC function. It means there is a delay introduced by a BCP which must make an evaluation of BCH and set the OXC which also takes a time. Current OXCs take a micro seconds to switch. There are articles proposing principles allowing the switching in scale of nanoseconds and in near future these principles are going to be used in OXCs on the market. The velocity of light in a fibre is not 3e8*mps* but less. For example a 200 *km* long fibre introduces delay 1 ms of signal. 1ms propagation delay is higher than switch time of newest switches. Thus all these delays must be carefully considered and introduced into all time domain calculations.

The paper is organised as follows: Chapter 2 introduces a problem of routing and wavelength assignment in labelled OBS. There are two paragraphs sorting out the problem of

routing and wavelength assignment. There are provided formulas for a proper time offset evaluation and criterion of shortest path. Chapter 3 considers effects of inappropriate time evaluation with respect to modern optical modulation formats and time of OXC switch. Chapter 4 concludes the article and suggests future research on topic of labelled OBS.

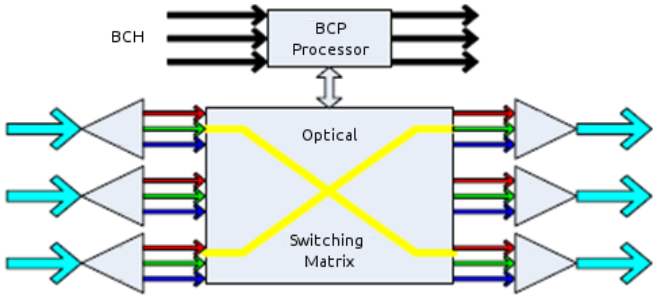


Figure 1. Fundamental parts of labeled OBS CoreNode.

II. CURRENT TESTBEDS

Proper time offset evaluation and research of impairment aware routing needs an optimal simulation models or real world OBS switch which is very expensive and its use is limited. For OMNeT++ simulator only one library exists implementing labelled OBS [2] but this library is only for OMNeT++ version 3. The library is not fully compatible with OMNeT++ version 4. The code does not support dynamic routing algorithms, impairments are not taken into account and only static time offset can be set. In spite of claims on project's website this library is not evolving further, in best knowledge of authors. Additional research cannot be done without such models so models must have been re-implemented with reusability and well documentation in mind to allow also other researchers to carry on in a research.

The main problem of an all-optical network library is how to process a data bursts in optical domain. This is treated differently for OPS, OCS and OBS but a similar problem of a lightpath establishment remains. The problem can be broken in two parts, routing and wavelength assignment (RWA). These cannot be solved at one time because the RWA problem is considered NP problem [3]. In RWA the routing must be solved first and then wavelengths assignment along the path [3]. The wavelength conversion (WC) and polarization multiplex (PM) [4] avoids wavelength assignment constraint and decreases blocking probability reducing a probability of successful data transfer. To have fully functional all-optical OMNeT++ library of labelled OBS, WC and PM are introduced as well.

A. Path evaluation

The process of finding an optical lightpath relies on knowledge of a network topology that can be described as *G(N,E)* where N stands for nodes and E stands for edges or links between two adjacent nodes. Links involved into the path must fit to criterion given by Bellman equation (1) which says that shortest path between two endsystems can be found recursively testing (1) every incoming link.

(1)

Where *i, j* stand for endsystem nodes in Figure 2, *u(i,j)* stands for a shortest path between node *i, j*. It must be mentioned that Bellman equation does not work for graphs with negative values of edges. Current implementation of (1) in the library uses distance between two adjacent CoreNodes as metric *a(k,j)*. Impairment aware extension of metric *a(k,j)* may introduce negative rating so the algorithm must be carefully checked in future.

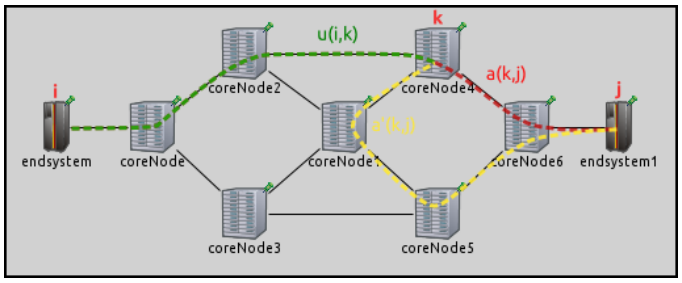


Figure 2. Measuring ready labeled OBS network.

A testbed containing CoreNodes and two Endsystem nodes. is depicted in Figure 2. Endsystem node is a compound model containing traffic generator, consumer and scheduler which implements mathematical models discussed in the next paragraph. Endsystem model's purpose is in evaluation of scheduling algorithms and reservation protocols. Tested algorithms are moved to EdgeNode model. This testbed was furthermore used for evaluation of time offset in next paragraph.

B. Time offset evaluation

The path obtained from previous paragraph serves for a routing purpose of a BHP that set up a lightpath out of set of available wavelengths of each CoreNode along the path. Once the lightpath is set up an optical burst might be sent into the lightpath. The just-enough-time (JET) signalling protocol is proposed for labelled OBS. JET is one way signalling protocol so there must be defined a delay when an optical burst can be sent onto the established lightpath [5]. The time offset between an optical burst and BHP must consist of the propagation delays introduced by fibres between CoreNodes along the path and delays introduced by CoreNodes. A time delay introduced by a fibre along a path can be calculated with (2) taking into account the length of a fibre *e* and fiber's core's index of refraction *n*.

(2)

For a long distance between two CoreNodes this propagation delay cannot be neglected and can be even higher than switch time of OXC. The OXC switch time vary from nanoseconds up to milliseconds depending of technology used. The propagation time of an optical burst along a lightpath is given by (3). Equation (3) is only valid for properly signalised lightpath. If a lightpath is not established a loss may occur.

(3)

Where $t(e_i)$ stands for a delay introduced by interconnecting fibre (2). The propagation time of a BCH is given by (4). BCH delay is affected by a distance between two feasible CoreNodes, i.e. (2), performance of AD/DA converters used for opto-electrical conversion for electrical BCP and by a performance of microprocessor used in BCP. The delay introduced by BCP microprocessor varies up to microseconds.

(4)

The time offset used in models is defined by (5). Equation (5) takes into account a number of nodes along a path, delay introduced by BCP, switch time of the same. The equation (5) comes from equation (4) and (5) subtraction and considerable switch time of the OXC.

(5)

Thanks to (5) the offset time can be easily formulated and any loss of data along a lightpath cannot occur. In case of violation of equation (5) a data loss may occur. The amount of lost data packets of an optical burst is further discussed in more details in next chapter.

III. IMPORTANCE OF CONSTRAINTS

Violation of equation (5) in sense of lower offset time can cause loss of some data packets in the burst or even a loss of a whole optical burst itself in extreme cases. The loss highly depends on an optical burst degree and the transmission speed not on the modulation format itself. In case of short switching time and burst arrival at the moment the OXC is not switched yet a part of an optical signal is lost. If a piece of an optical burst is lost it means a number of modulation symbols are lost. A number of lost modulation symbols depend on the modulation format i.e. its modulation speed. Through equation (6) can be obtained amount of user data lost in *Bits*.

(6)

Where v_{mod} stands for modulation speed in *Bd*, M stands for a number of modulation states, T_{OXC} stands for switch time of an OXC in *s* and v_{trans} stands for transmission speed in *bps*.

Table I. compares different modulation types and its performance against switching time which corresponds to a variety of switching matrix types. Two different types of phase modulation and one amplitude modulation format is there evaluated.

TABLE I. DATA LOSS COMPARISON OF DIFFERENT MODULATION FORMATS AND DIFFERENT SWITCHING TECHNOLOGIES

Data loss [B]	**OXC switch time [µs]**		
	0.010	***10***	***10,000***
2.5G OOK	3.13	3125.00	3.13e6
10G BSPK	12.50	12500.00	12.50e6
40G BSPK	50.00	50e3	50e6
40G QSPK	50.00	50e3	50e6

Amount of lost data depends on OXC switch time T_{OXC} amount is obtained through equation (6). From Table I. is clear the demand of high speed optical switches. Today accessible systems with transmission speed 10 Gbps may suffer with loss of 12.50B in case of inaccurate time offset set up. Such a loss causes destruction of a first IP packet's header of harmed optical burst. Current on market switching matrices might cause total loss of a number of IP packets. Such lost provokes a retransmission of incidental TCP datagrams which decreases total TCP performance [6]. Data loss and TCP retransmission in labelled OBS might be shrank only by OXC switch time reduction, it means with switching matrix technology evolution.

IV. CONCLUSION

Article described a RWA problem in labeled OBS networks. Along the routing problem is there proposed formula (5) for accurate time offset evaluation of optical burst overhead in front of an optical burst. Must be mentioned equation (5) is not exact but on the other hand fully satisfies demand of time offset. Difference between analytical equation (5) and precise approximation carried out from simulations is negligible.

In all calculations values of BCP, OXC delay are supposed to be know but in real world implementation it is not possible, especially in multi-vendor network, to assume a value a delay. To spread such information is needed an extension of signalling protocol. This can be satisfied through GMPLS plain or an arbitrary protocol using messages similar to BHP carried through a separate wavelength. Information about fibre's impairments can be spread this way as well. Then classical RWA problem turns into impairment aware RWA (IA-RWA). Bellman equation can be further used but must bear in mind to prepare relevant criterion which does not create a negative value of an edge or link. This problem is very promising and very important for further deployment of all types of all-optical systems.

ACKNOWLEDGMENT

The work was supported by grant The Research in the Area of the Prospective Information and Communication Technologies under project MSM6840770014 and by grant Development and Optimization of the Simulation Modules for Communication Networks under project SGS11/124/OHK3/2T/13.

REFERENCES

[1] RUMLEY, S, et al. Towards Digital Optical Networks. Germany : Springer, 2009. Software Tools and Methods for Research and Education in Optical Networks, s. 362. ISBN 0302-9743.

[2] https://www.tlm.unavarra.es/investigacion/proyectos/strrong/soft/obsmodules/

[3] BINH, L.N.; CIEUTAT, Charles. Routing and Wavelength Assignment in GMPLS-based Optical Networks : An OMNeT++ modelling platform. In Routing and Wavelength Assignment in GMPLS-based Optical Networks: An OMNeT++ modelling platform. Monash University : Monash University, 2003. s. 9

[4] KOZÁK, Miloš; BOHÁČ, Leoš. Knowledge in Telecommunication Technologies and Optics - KTTO 2010. Ostrava : VŠB - TUO, FEI, Katedra elektroniky a telekomunikační techniky, 2010. The Optical Burst Switching OMNET++ Simulator Model Extension with The Polarisation Multiplex, s. 36-40. ISBN 978-80-248-2330-0.

[5] DOLZER, Klaus, et al. Evaluation of Reservation Mechanisms for Optical Burst Switching. AEÜ - International Journal of Electronics and Communications [online]. 2001, vol. 55, no. 1 [cit. 2009-06-30], s. 1-8. Dostupný z WWW: <http://www.ikr.uni-stuttgart.de/Content/Publications/Archive/Dz_Aeue-final01_33136.pdf>.

[6] VENKATESH, T; SIVA RAM MURTHY, C. An Analytical Approach to Optical Burst Switched Networks. London : Springer, 2010. 277 s. ISBN 978-1-4419-1509-2

Miloš Kozák received the M.S. degree in electrical engineering from the Czech Technical University, Prague, in 2009. He participates in teaching of course of optical communication systems. His research interest is on the application of high-speed optical transmission systems in a data network. Currently he is actively involved in some of the projects on high speed optical modulations and Communication of Czech Antarctic Station.

Leoš Boháč received the M.S. and Ph.D. degrees in electrical engineering from the Czech Technical University, Prague, in 1992 and 2001, respectively. Since 1992, he has been teaching optical communication systems and data networks with the Czech Technical University, Prague. His research interest is on the application of high-speed optical transmission systems in a data network. He has also participated in the optical research project CESNET - the academic data network provider to help implement a long-haul high-speed optical research network. Currently he is actively involved in and leads some of the projects on optimal protocol design, routing, high speed optical modulations and Turbo codes for future data and optical networks.

Noise analysis of dark rooms and their evaluation

Stanislav Hejduk, Vladimir Vasinek, Jiri Bocheza, Jan Vitasek, Jan Latal, Artem Ganiyev
Faculty of Electrical Engineering and Computer Science, Department of Telecommunications
VŠB-Technical University of Ostrava
17. listopadu 15/2172, 708 33 Ostrava Poruba, Czech Republic

Abstract— **In this article we will deal with optical noise inside dark room. From the perspective of noise power in the room and from the perspective of statistical noise distribution. There are two dark rooms, where we measure lighting conditions. Because the parameters of the rooms vary greatly, so in the end we try to define the conditions of room construction to achieve the lowest noise. We also try to define what should be avoided in actual measurements.**

Keywords; photon counter; cps; dark room; free space optics; fso; optical noise; optical noise distribution; poisson distribution

I. INTRODUCTION

With the development of indoor free space optical communications (FSO) it emerged the need to measure the distribution of light in free space. For purposes of this kind of measurement it is necessary to eliminate ambient light, which is also a source of noise. Without elimination of ambient light, measured values would be affected by the significant error.

In the design phase we can be satisfied with software simulations [1], but sooner or later will be required experimental verification. For these measurements we have two laboratories where we can perform measurements in a dark environment.

Our task was to measure and analyze the noise properties of these laboratories. It means that we will not only interest only in the average value of measured performance, but we will also interest in shape of signal, because these values change over time and this is a source of noise.

Knowledge of the noise distribution is especially important for determining the accuracy of measurement and specification of the uncertainties in determining the distribution of optical power in space and SNR evaluation. If the measured optical power didn't change over time, we could simply measure average power without useful signal and then subtract it from the measured values. In case of high-power signals measurement the background noise isn't usually very significant. Nevertheless the future of FSO aims to reduce power due to energy consumption, which limits the use of this technology in mobile devices. So the low-power signals measurement is necessary.

Our goal will be to determine which values can be expected in our dark labs. Low-light conditions in both labs exclude to use of standard photo-detector instruments. To measure under these low-light conditions we have PHOTON COUNTER PerkinElmer type MP 984 RS. So we are able to measure light conditions in the room up to the level of individual photons.

This article was created with the active support of the Czech Science Foundation GA102/09/0550 and SP2011/47.

II. PREPARATIONS

A. Description of examined laboratories

The first examined laboratory (lab1) is fully equipped optical lab, so we can find a lot of equipment inside, like computer and ceiling fluorescent lighting. For darkening there are used black curtains around the perimeter of the room and the room has no windows. The floor and ceiling of the room are not obscured. The floor is paved, the surface of tiles is matt and ceiling is white.

The other laboratory (lab2) was built especially for needs of darkening. It's in fact a room inside a room that is inside completely painted with matte black paint and in the room is nothing, what could be a source of light.

B. Test conditions

The basic precondition was that the noise measurements were made under normal operating conditions, which means that the measurements were carried out in daylight. Theoretically, we wouldn't be interest in ambient environmental conditions, but doors of both rooms caused a big problem, because even the tiniest leak allowed the penetration of ambient light into the room. It caused a significant increase in the measured values of optical noise. During the measurements in lab1 there were also turned off all the devices that weren't necessary for measurement.

The measurement itself was carried out until the number of readings didn't exceed 1000 values. This number can be considered as sufficient for statistical analysis of measured values.

C. Format of measured data

Results are reported in CPS units (Counts Per Second) and represent the number of photons recorded by photon counter in one second period. Measurements in very dark rooms are performed by counting photons for some time (for example 5seconds) and then calculate the average value over a period of one second. So the measured values can be expressed also in decimals. But we must take into consideration the fact that photon is a particle, which can't be divided into smaller parts.

III. THEORY

Measurements of extremely low power means that background noise levels will be evaluated on level of individual photons. The dominant noise in this case is quantum noise. It has its origin in the quantum nature of light. Quantum behavior of electromagnetic radiation must be taken into consideration at

optical frequencies, when hf > kT and quantum noise dominates over thermal fluctuations.

Photodiode light detection is a discrete process, because the creation of electron-hole pairs resulting from photon absorption. The signal that emerges from the photo detector, subjects to the statistics of incoming photons. Depending on the type of incident light it can be distinguished coherent and incoherent radiation. For the incident number of photons, which are statistically independent when impact on the detector applies that they subject to discrete probability distribution. This distribution is independent of the previous number of detected photons. It was found [2] that the probability P (n), that number of photons will be detected in time interval τ, where n* is the expected average number of photons, is described by Poisson distribution [3]. Then applies the relation:

$$P(n) = \frac{e^{-n^*} \cdot (n^*)^n}{n!} \quad (1).$$

n* is simultaneously the scattering distribution. For the Poisson distribution is typical that the mean and variance are the same size. If it wasn't measured in the number of photons, but optical power, the relationship between the optical power and the mean number of photons incident would be given as follows.

$$n^* = \frac{P_0 \cdot \tau}{\frac{h \cdot c}{\lambda}} \quad (2)$$

Incoherent light is emitted by independent atoms, and therefore there isn't any relation between the emitted photons. Then the distribution of photons corresponds to an exponential distribution, which is averaged in the Poisson distribution. Which is described as:

$$P(\mathrm{n}) = \frac{n^*}{(1+n^*)^{(n+1)}} \quad (3).$$

The following two figures shows the theoretical curves of the Poisson distribution for the mean values measured in both rooms.

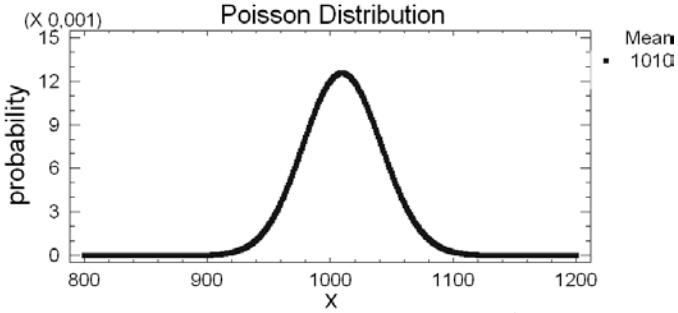


Figure 1. Theoretical Poisson distribution curve for lab1.

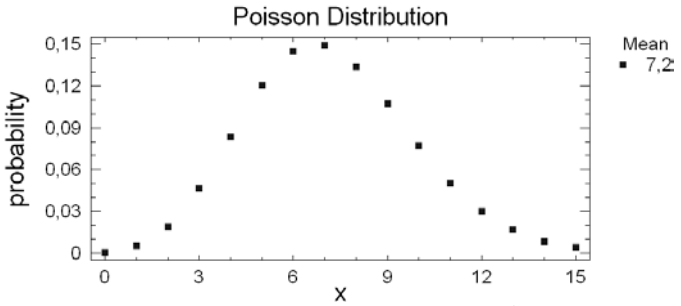


Figure 2. Theoretical Poisson distribution curve for lab2.

In conclusion, these theoretical values were compared with actual measured values.

IV. RESULTS

A. *Measurement in lab1*

The first room, which was measured, was lab1. In Figure 3 we can see a graphical representation of the number of photons per second measured in the timeline.

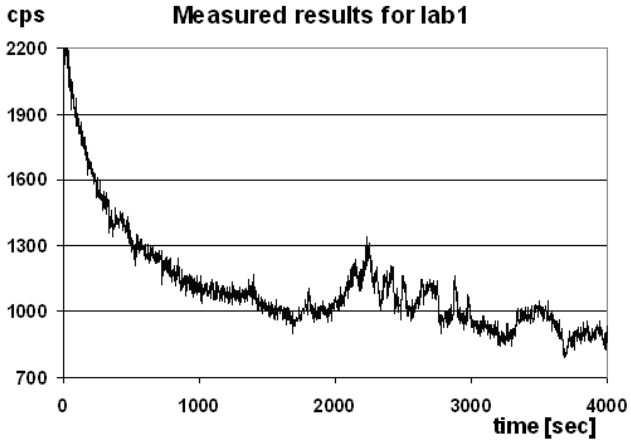


Figure 3. The values measured in lab1.

From the Measured values in Figure 3 we can see how the fluorescent lamps, during its cooling, influence the light conditions in the laboratory. For the following analysis, values affected by the fluorescent lamps cooling time will be ignored, because they would too much distort the results. So the first 20 minutes will not be considered for further calculations, but for measurements in the lab we can't ignore this problem. Other fluctuations in the remaining measured values can be attributed to changes in ambient room lighting situation due to changing weather and movement of people outside the doors. These changes can be eliminated by careful sealing around the doors, otherwise the results will be burdened by a significant error.

From the measured data, we can easily determine the average value of photons per second.

CPS_{AVG}=1009,98 [photons/second]

To express the value of the CPS in watts we can use the following formula:

$$P = \frac{h \cdot c}{\lambda} \cdot CPS \text{ [W]} \quad (4).$$

For the conversion we need to know the value of λ, because the photon energy is different for different wavelengths. Because we don't know the value of λ for every individual photon, the converted values will be only approximated. Assuming that the used photoncounter is sensitive to wavelengths between 300 - 650nm, then after substitution into equation (4) we get values ranging from $3.1*10^{-16}$ to $6.7*10^{-16}$ W. The average power will thus be

$$P_{AVG} = 5 \cdot 10^{-16}\ W\ .$$

For further analysis of the noise distribution we have to look at the graph on Figure 4, which shows the distribution of CPS values in measured sample. Their comparison with the estimated value is then shown in Figure 5, where we can see a histogram of the measured values interleaved with ideal Poisson distribution.

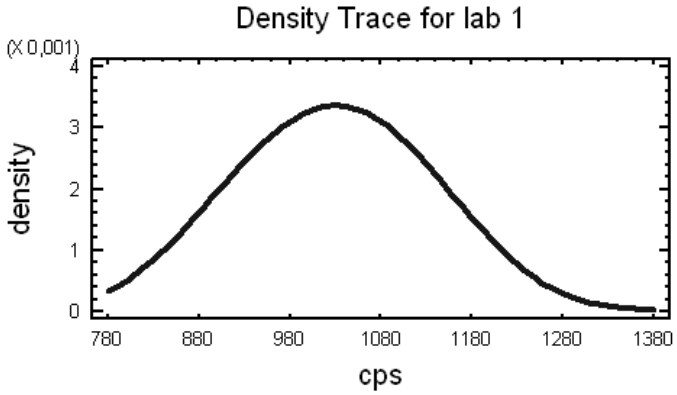


Figure 4. Density trace for values measured in lab1.

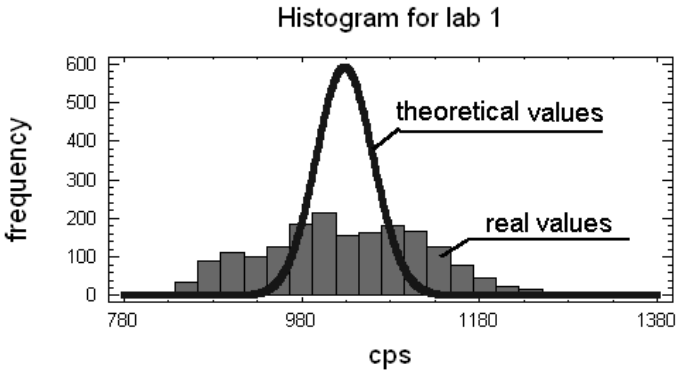


Figure 5. Comparison of measured values with the theoretical course.

If we try to mitigate the impact the neighborhood and use only the linear part of the measured values, from 850-1400 seconds of time, graph from Figure 5 will change to the course of Figure 6. Now we can clearly see how the theoretical and measured values overlap each other. Unfortunately for conclusive statistical analysis of this part of the course we don't have enough data (at least 1000 values are necessary).

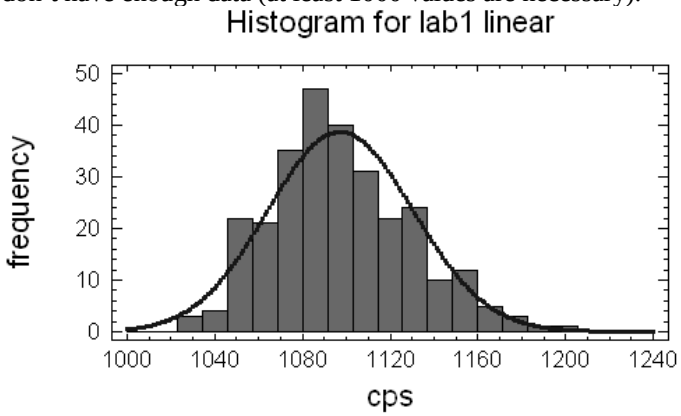


Figure 6. Comparison for the linear part of the measured values.

B. Measurement in lab2

If we look at the values measured in the other laboratory shown in Figure 7, at the first sight the magnitude of measured values are several orders lower.

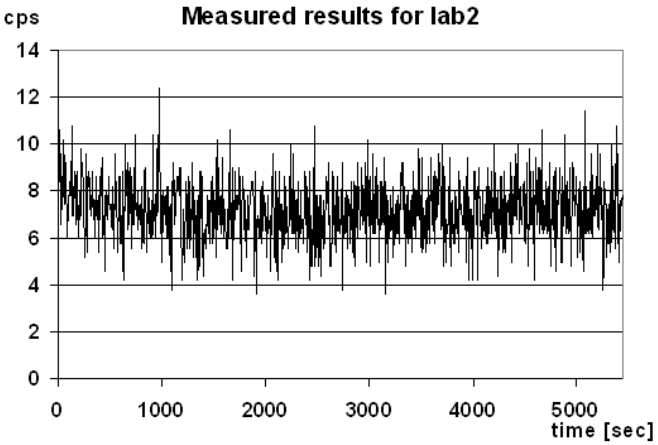


Figure 7. The values measured in lab2.

Despite the fact that the measurements took place during the day, the measured values was only in units of photons, which ensures very low optical noise in this room. The average value of the CPS in this case was very low.

CPS_{AVG} = 7,2 [photons/second].

For conversion of CPS units to optical power, we use equation (4), as in previous case. After substitution we can calculate an approximate value of power.

$$P_{AVG} = 3{,}45 \cdot 10^{-18}\ W\ .$$

If we compare these values with the lab1, we can clearly say that the lab2 achieves much better parameters.

For a closer look to the noise characteristics we have to look at figure 8, which shows the distribution of measured values.

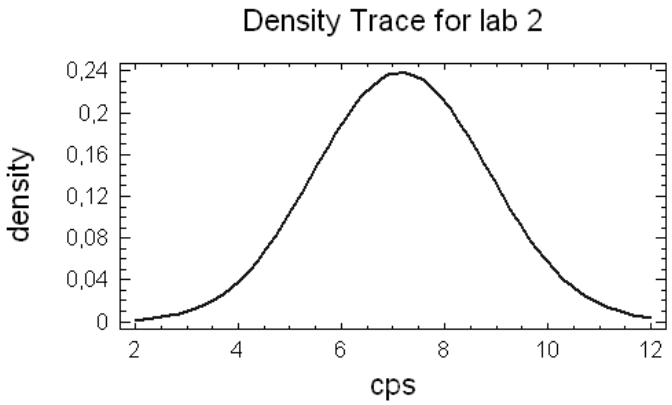


Figure 8. Density trace for values measured in lab2.

Figure 9 shows histogram of the measured values interleaved with theoretical Poisson distribution. As we can see, the measured values in this room are even better than the theoretical premise.

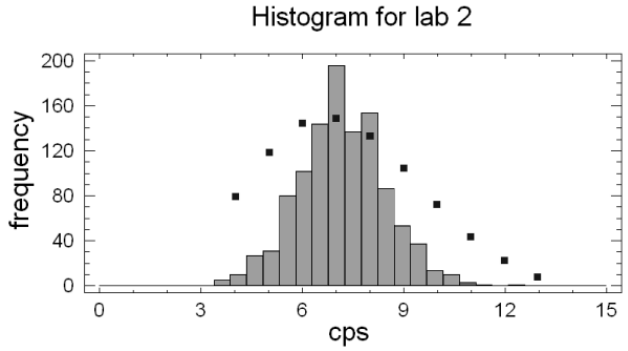


Figure 9. Comparison of measured values with the theoretical course.

C. *Photon counter characteristics*

In the data analysis, we must not forget the error of photon counter itself. Error unit is described as a darkcount value, which represents the number of photons detected by the photodetector in absolutely dark environment. Number of darkcounts prescribed by the manufacturer is 1CPS. Value achieved during the verification was 0.71CPS, which corresponds to the reported value. This value may seem like insignificant, but for the values measured in the lab2 it has represents an error that is worth mentioning.

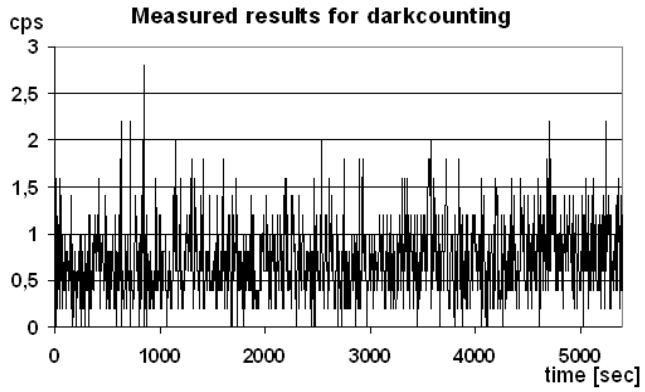


Figure 10. Measured darkcounts.

The measured values are shown in figures 10 and 11. These values are also quite close to values measured in the lab2 at night conditions.

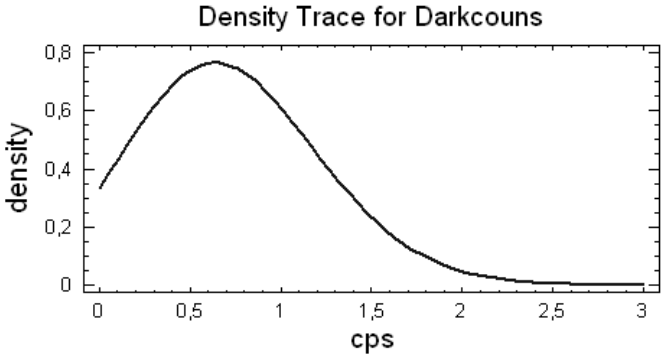


Figure 11. Density trace of the measured darkcounts.

D. *Summary of the measured results*

The following table (TAB1) compares the measurement results for both laboratories and darkcounts of photon counter.

TABLE I. MEASURED RESULTS

Measured event	CPS_{AVG}	P_{AVG} [W]
Lab 1.	1009,98	$5 \cdot 10^{-16}$
Lab 2.	7,17945	$3{,}45 \cdot 10^{-18}$
Darkcounts	0,715838	$4{,}84 \cdot 10^{-19}$

E. *Construction and measurement rules*

Comparing the measurement results can be derived some rules for construction and measurement in dark rooms.

- Prevent the penetration of ambient light into the room.
- Eliminate all light sources that are inside the room.
- Remove all fluorescent materials from the room.
- Mitigate impact of multipath propagation of light [4] inside the room (use the dark walls).

After the room preparation, we are confronted with another problem, which is nothing more than a way to measure inside the room. The best is to ensure that we can subtract the data from outside the room. Otherwise we have to find a way how to read the measured data in terms of a dark room. This way can be difficult, because each light source is also source of noise.

V. CONCLUSION

The objective of measurements in a dark room is to maximize the signal to noise ratio, so the primary task is the best darkening that can be achieved. Ideally the room is empty during measurements and light does not penetrate into it from outside. However, if the measurement requires the presence of a person or a device inside the room, it's necessary to confine the light emitting elements. For example, fluorescent watch hands emitted so many photons that one measurement had to be completely repeated because the results were unusable.

Regarding the statistical viewpoint, the aim was to compare theoretical waveforms of Poisson distribution with measured values. Taking into account the variations, we can say after comparison of the curves that the distribution of optical noise corresponds to the Poisson distribution and can therefore be described by the relations reported in the theoretical part. Poisson distribution is asymmetric, so the variance of mean value is not same on both sides. This rule corresponds to all measurements shown in Figures 4 and 8 as well as theoretical results in Figures 1 and 2.

So if we compare our laboratories, we can say that the lab1 has a worse performance than the second laboratory. Noise in the lab1 is due to significant penetration of ambient light even worse than the predicted theoretical value. On the contrary, the value of optical noise in the lab2 was slightly better than the expected theoretical value.

REFERENCES

[1] Jaw-Luen Tang, Yao-Wen Chang, Design of an Omnidirectional Multibeam Transmitter for High-Speed Indoor Wireless Communications, EURASIP Journal on Wireless Communications and Networking, Volume 2010, Article ID 728468.

[2] Saleh,B.E.A., Teich,M.C.: Fundamentals of Photonics. Wiley and Sons,Inc. New Jersey, 2007. ISBN: 978-0-471-35832-9.T. Dostál, V. Axman, Elektrické filtry, FEKT VUT, Brno, 2002.

[3] Kolektiv autorů: Aplikovaná matematika – oborová encyklopedie. SNTL Praha 1978.

[4] Adrian Mihaescu, Pascal Besnard: Indoor Wireless Optical Communications, 8th International Conference on Communications, Bucharest, Romania, 2010.

Time varying impulse response of a PMD affected single mode fibre optic link with polarization scramblers

Zbigniew Lach
Faculty of Electrical Engineering and Computer Science, Lublin University of Technology
Lublin, Poland
z.lach@pollub.pl

***Abstract*—Optical fibre line with polarization scramblers is modelled as a time-variant linear system. The formula for time varying impulse response of such a system is derived and relations to PMD and to parameters of scramblers control are shown. One possible application of the time varying impulse response is considered and illustrated with results of simulations.**

Keywords- polarization mode dispersion, single mode fibre, polarization scrambler

I. INTRODUCTION

Optical fibers suffer from polarization mode dispersion (PMD) which affects capability of links using using On-Off-Keying (OOK) signalling to achieve transmission capacity exceeding 10Gbps per wavelenght channel [1]. A number of methods were developed to combat the effects of PMD [2]. One case of practical interest is the link with burried or undersea cables in which temporal variations of PMD are relatively slow i.e. PMD can be static ("frozen") for days or even months [3]. Slow changes of PMD result in long lasting link outages. For a link of such type the technique using polarization scrambling distributed along the link span in combination with proper Forward Error Correction (FEC) was shown a method of choice [2]. In this method "frozen" PMD becomes time-varying which potentially shortens time intervals of unacceptably high signal deterioration. Short error bursts can be corrected with the use of an appropriate FEC which potentially converts an inoperable link into a fully functional low bit error rate (BER) one.

Theoretically, this technique is effective provided each polarization scrambler uniformly distributes it's output states of polarization (SOP) on the Poincare sphere within each data transmission frame [4]. This would require scrambling with unrealistic infinite speed. In practical realizations a question arises on how to control the scramblers with real world signals to satisfy the conditions that are imposed by the ability of the FEC to correct error bursts. One way of analysing this aspect can be to inspect the time varying impulse response (TVIR) of the polarization scrambled fibre optic link. In this paper such idea is developed and prospects for its use are shown through examples. Analytic model which describes the TVIR of a fibre optic polarization scrambled line as a function of parameters of fibre segments, of control signals, etc. can be useful in design of links with PMD affected fibres as well as computer simulations.

II. LINEAR TIME-VARIANT MODEL OF OPTICAL FIBRE LINE WITH POLARIZATION SCRAMBLERS

Generally, silica fibres are nonlinear media however, under certain conditions assumption of approximate linearity holds [5] which allows to model such a fibre as a linear system. Such an assumption will be made for the following text.

A single mode fibre supports two polarization modes. The input and output of the system which is the model of a single mode fibre can be considered 2-dimensional vectors with components describing the polarization orthogonal constituents of light. Optical signals propagating in fibre optic telecommunication links can be regarded modulated monochromatic polarized waves [6]. Such a wave can be described by 2-dimensional complex envelope vector $\hat{\mathbf{s}}(t)$ defined by:

$$\hat{\mathbf{s}}(t)=\begin{bmatrix}\hat{s}_x(t)\\ \hat{s}_y(t)\end{bmatrix}=\begin{bmatrix}a_x(t)e^{\varphi_x}\\ a_y(t)e^{\varphi_y}\end{bmatrix}, \tag{1}$$

which is related to the wave $\mathbf{s}(t)$ by:

$$\mathbf{s}(t)=\begin{bmatrix}s_x(t)\\ s_y(t)\end{bmatrix}=\mathrm{Re}\left\{\hat{\mathbf{s}}(t)e^{j\omega_0 t}\right\}. \tag{2}$$

Here, $s_x(t)$ and $s_y(t)$ are polarization components of the wave $\mathbf{s}(t)$ while pairs $(a_x(t),\ \varphi_x(t))$ and $(a_y(t),\ \varphi_y(t))$ represent their envelopes and phases. The ω_0 is the angular frequency of the optical carrier.

The paper analyses the case in which polarization scramblers are inserted at discrete points of a fibre link in order to animate PMD. Time varying polarization mixing occurs in fibres in the effect of scrambling. Thus, TVIR of such a link shows temporal variations and the link can be regarded a time variant linear system. Its' output signal is related to the input signal via [7]:

$$\mathbf{w}(t)=\int_{-\infty}^{\infty}\mathbf{h}(t,t-t_2)\mathbf{v}(t_2)dt_2=(\mathbf{h}\bullet\mathbf{v})(t). \quad (3)$$

In (3) $\mathbf{v}(t)$ and $\mathbf{w}(t)$ are 2-dimensional signals of the form described by (2) which correspond to the input and output waves, respectively. The $\mathbf{h}(t,t_2)$ is the time varying impulse response (TVIR) of the system:

$$\mathbf{h}(t,t_2)=\begin{bmatrix} h_{xx}(t,t_2) & h_{yx}(t,t_2)\\ h_{yx}(t,t_2) & h_{yy}(t,t_2)\end{bmatrix}. \quad (4)$$

The scalar components $h_{ij}(t,t_2)$ represent in time domain the transfer characteristics of the system with respect to the two polarization modes. The t parameter is the "running time" characterizing the time flow in the system while the t_2 parameter is the "lag", the delay with respect to the running time.

In general, the components of $\mathbf{h}(t,t_2)$ take account for all factors affecting propagation of polarization modes: polarization mode dispersion, other types of dispersion (polarization independent), polarization dependent loss (PDL), and time varying polarization mixing due to scrambling.

Considering that the input and output signals of the system of interest are modulated sinusoids it is more convenient to express the input-output relation in terms of their complex envelopes. The relation equivalent to (3) is given by [7]:

$$\hat{\mathbf{w}}(t)=\int_{-\infty}^{\infty}\hat{\mathbf{h}}(t,t-t_2)\hat{\mathbf{v}}(t_2)dt_2=(\hat{\mathbf{h}}\bullet\hat{\mathbf{v}})(t), \quad (5)$$

where: $\hat{\mathbf{v}}(t)$ and $\hat{\mathbf{w}}(t)$ are the complex envelopes (as defined by (1)) of the input and output waves, respectively. The $\hat{\mathbf{h}}(t,t_2)$ is the "baseband" TVIR of the system defined as follows:

$$\hat{\mathbf{h}}(t,t_2)=\mathbf{h}(t,t_2)\exp(-j\omega_0 t_2). \quad (6)$$

It will be further referred to as the complex time variant impulse response (CTVIR). One shall note that with the definition (6) the function $\mathbf{h}=\mathbf{h}(t=const,t_2)$ need not be bandpass which is the case of any PMD model.

For a system that is a concatenation of a number of time variant linear systems the effective CTVIR shall be a linear transformation of the CTVIRs of the component systems. Let's limit, for the moment, to a cascade of only two systems, each with the corresponding CTVIR $\hat{\mathbf{h}}_1(t,t_2)$ and $\hat{\mathbf{h}}_2(t,t_2)$ (the systems are indexed starting from the input in the direction to the output of the cascade). Acording to (5) the resultant input-output relation of the cascade of the two systems is given as follows:

$$\hat{\mathbf{w}}(t)=(\hat{\mathbf{h}}_2\bullet(\hat{\mathbf{h}}_1\bullet\hat{\mathbf{v}}))(t). \quad (7)$$

The CTVIR $\hat{\mathbf{h}}_e(t,t_2)$ of the compound system can be derived from (7) after $\hat{\mathbf{v}}(t)$ is replaced with $\boldsymbol{\delta}(t-t_2)=[\delta(t-t_2),\delta(t-t_2)]^T$. This leads to the following formula:

$$\hat{\mathbf{h}}_e(t,t_2)=(\hat{\mathbf{h}}_2\circ\hat{\mathbf{h}}_1)(t,t_2)=\int_{-\infty}^{\infty}\hat{\mathbf{h}}_2(t,\xi)\hat{\mathbf{h}}_1(t-\xi,t_2-\xi)d\xi. \quad (8)$$

The operator $\circ$ appearing in (8) denotes the time-variant convolution generalized to two-dimensional signals which is formally defined by the last expression in (8).

Generalizing the result in (8) to the case of a cascade of any M components one obtains the effective CTVIR:

$$\hat{\mathbf{h}}_e(t,t_2)=(\hat{\mathbf{h}}_M\circ\hat{\mathbf{h}}_{M-1}\circ\ldots\circ\hat{\mathbf{h}}_1)(t,t_2), \quad (9)$$

where $\hat{\mathbf{h}}_M(t,t_2)$, $\hat{\mathbf{h}}_{M-1}(t,t_2)$,..., $\hat{\mathbf{h}}_1(t,t_2)$ are the CTVIRs of components indexed from the cascade's input to its' output.

III. Complex Time-Variant Impulse Response of a Fibre Optic Line with Polarization Scramblers

A. Complex time-varying impulse response of a cascade of DGD components

Let's start the derivation from an example being a cascade of time-invariant components, each with fixed DGD. The well known 2×2 Jones matrix representing a frequency domain transfer function of a single component of this type can be rearranged to the form:

$$\mathbf{U}(\omega)=e^{-j0.5\omega\tau}\mathbf{U}_0+e^{j0.5\omega\tau}\mathbf{U}_1, \quad (10)$$

where: $\mathbf{U}_0$ and $\mathbf{U}_1$ are frequency invariant matrices with elements determined by the principal states of polarization (PSP) and PDL, while τ is the DGD of the component. In general, the $\mathbf{U}_0$ and $\mathbf{U}_1$ are given by:

$$\begin{aligned}\mathbf{U}_0&=\cos\alpha\begin{bmatrix}\cos\alpha & \sin\alpha\\ \sin\alpha & \cos\alpha\end{bmatrix}10^{-0.05\Delta}\\ \mathbf{U}_1&=\sin\alpha\begin{bmatrix}\sin\alpha & -\cos\alpha\\ -\cos\alpha & \sin\alpha\end{bmatrix}10^{0.05\Delta}\end{aligned}, \quad (11)$$

where: α is a rotation angle of the component's birefringence axes with respect to the $X0Y$ axes and Δ is the PDL of the component in decibels. The corresponding time-invariant impulse response is as follows:

$$\mathbf{h}(t)=\delta\left(t-\frac{\tau}{2}\right)\mathbf{U}_0+\delta\left(t+\frac{\tau}{2}\right)\mathbf{U}_1. \quad (12)$$

The complex impulse response is obtained through multiplying (12) by $exp(-j\omega_0 t)$ which yields:

$$\hat{\mathbf{h}}(t) = \delta\left(t - \frac{\tau}{2}\right)\mathbf{U}_0 e^{-j\omega_0\frac{\tau}{2}} + \delta\left(t + \frac{\tau}{2}\right)\mathbf{U}_1 e^{j\omega_0\frac{\tau}{2}}. \qquad (13)$$

For a cascade of *N* fixed DGD components we get from (9) and (13):

$$\hat{\mathbf{h}}_e(t) = \sum_{i=0}^{2^N-1}\delta\left(t - \frac{\Delta\tau_i}{2}\right)\mathbf{U}_i \exp\left(-j\omega_0\frac{\Delta\tau_i}{2}\right), \qquad (14)$$

where:

$$\Delta\tau_i(t) = \sum_{n=0}^{N-1}(-1)^{d(N-n,i)}\tau_n(t) \qquad (15)$$

and:

$$\mathbf{U}_i = \prod_{n=0}^{N-1}\mathbf{U}_{N-n,d(N-n,i)}\,. \qquad (16)$$

In (14)-(15) above: τ_n is the fixed DGD of the *n*-th component in the cascade, *d(n,i)* is the value of the *n*-th digit of the *N*-bit binary representation of the index *i*, and $\mathbf{U}_{n,0}$ and $\mathbf{U}_{n,1}$ are the $\mathbf{U}_0$ and $\mathbf{U}_1$ of the *n*-th component, respectively.

The result in (16) can be easily generalized to a time-variant case when all DGDs in the cascade are allowed to vary with time. From (14) one gets the following expression for the CTVIR of a cascade of time-variant DGD components:

$$\hat{\mathbf{h}}_e(t,t_2) = \sum_{i=0}^{2^N-1}\delta\left(t_2 - \frac{\Delta\tau_i(t)}{2}\right)\mathbf{U}_i \exp\left(-j\omega_0\frac{\Delta\tau_i(t)}{2}\right). \qquad (17)$$

The first term in (17) represents time shifts which the complex envelope undergoes due propagation through the cascade while the second one describes the associated phase shifts of the optical carrier.

B. Complex time-varying impulse response of a polarization scrambler

There may be a variety of realizations of polarization scramblers. In order to rotate polarization of light one of following two components can be utilised:

- a space rotatable waveplate or,
- a tunable phase shifter,

typically stacked in a cascade. Polarization scramblers build with the two types of components are equivalent. In the following only the polarization scrambler based on tunable phase shifters will be considered without loss of generality.

A three axis scrambler can be build as a cascade of three variable DGD components playing the role of phase shifters (at a given optical frequency DGD changes manifest as phase shifts to a sine wave). Two of them shall have the same orientation of their birefringence axes, say 0 rad with respect to XOY axes, and the third shall be sandwiched between the two and shall have birefringence axes rotated by $\frac{1}{4}\pi$. This rule is kept observed while adding extra stages to a scrambler. One conclusion from the above considerations is that a polarization scrambler, being a cascade of variable DGD components, has its CTVIR in general described by (17).

In the following the (17) will be specialized to express a CTVIR of an *N*-stage scrambler (a stage is regarded a single phase shifter). For a phase shifter with birefringence axes aligned with *X0Y* axes the matrices $\mathbf{U}_0$ and $\mathbf{U}_1$ are as follows:

$$\mathbf{U}_0^s = \begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix} \qquad \mathbf{U}_1^s = \begin{bmatrix}0 & 0\\ 0 & 1\end{bmatrix}, \qquad (18)$$

while for a phase shifter with birefringence axes rotated by $\frac{1}{4}\pi$ the matrices are given by:

$$\mathbf{U}_0^s = 0.5\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix} \qquad \mathbf{U}_1^s = 0.5\begin{bmatrix}1 & -1\\ -1 & 1\end{bmatrix}. \qquad (19)$$

In (18) and (19) above the upper script *'s'* lables matrices that correspond to the scrambler.

Moreover, it shall be noted that for any practical scrambler the $|\Delta\tau(t)|$ remains in the range of 10^{-2} picoseconds which is two orders of magnitude lower than DGD of a typical fibre segment. Consequently, in practical applications to an excellent approximation the polarization scrambling can be modelled with neglected effect of time shifting the signal's complex envelope:

$$\hat{\mathbf{h}}_s(t,t_2) = \delta(t_2)\mathbf{S}(t) \qquad (20)$$

where $\mathbf{S}(t)$ represents phase shifts due to polarization scrambling.:

$$\mathbf{S}(t) = \sum_{i=0}^{2^N-1}\mathbf{U}_i \exp\left(-j0.5\omega_0\Delta\tau_i(t)\right). \qquad (21)$$

In typical applications scramblers are driven by sinusoidal signals (particularly when fast scrambling, in the range of Megahertz, is applied). For an *N*-stage scrambler controlled by sinusoidal signals we have:

$$\mathbf{S}(t) = \sum_{l=0}^{2^N-1}\left\{\mathbf{U}_i \exp\left[-j\sum_{n=0}^{N-1}(-1)^{d(n,l)}\,j\Delta\varphi_n \sin\left(\omega_n t + \vartheta_n\right)\right]\right\}. \qquad (22)$$

In the above formula ω_n, ϑ_n are the angular frequency and the initial phase of the sinusoidal signal controlling the *n*-the stage of the scrambler and the $\Delta\varphi_n$ is the amplitude of this control signal; $\Delta\varphi_n = \pi$ for odd *n* and $\Delta\varphi_n = \frac{1}{2}\pi$ for even *n*.

C. Complex impulse response of a fibre segment

Here, PMD in any fibre segment constituting a fibre line with polarization scramblers is considered time-invariant, at least in the time scale of interest. Dynamic behaviour of such a

fibre segment is described by its time-invariant complex impulse response. Limiting to the case in which modelling PMD in a fibre by the first order model (frequency invariant DGD) is sufficient, the frequency domain transfer function of a fibre segment as well as its complex impulse response function can be described by (10) and (14), respectively. The matrices $\mathbf{U}_0$ and $\mathbf{U}_1$ are given by (11). Any fibre segment is fully characterized by its DGD (denoted as τ) and the angle of rotation of its birefringence axes (denoted as α).

D. Complex time-varying impulse response of a fibre optic line with distributed polarization scramblers

With the use of (9), (13) and (20) the CTVIR of a concatenation of a polarization scrambler followed by a fibre segment can be derived as:

$$\hat{\mathbf{h}}_{fs}(t,t_2)=\delta\left(t_2-\frac{\tau}{2}\right)\mathbf{H}_0(t)e^{-j\omega_0\frac{\tau}{2}}+\delta\left(t_2+\frac{\tau}{2}\right)\mathbf{H}_1(t)e^{j\omega_0\frac{\tau}{2}}, \quad (23)$$

where:

$$\mathbf{H}_0(t)=\mathbf{U}_0^f\mathbf{S}(t) \qquad \mathbf{H}_1(t)=\mathbf{U}_1^f\mathbf{S}(t). \quad (24)$$

The upper script '*f*' is introduced to differentiate the matrices $\mathbf{U}_0$ and $\mathbf{U}_1$ from the ones of a polarization scrambler.

Because (23) has the same form as (13) one can use the same reasoning that was used in the sub-section A above to draw the conclusion that the formula for the CTVIR for a cascade of *M* fibre segments, each preceded with a polarization scrambler, has the general form of (14) with time varying $\mathbf{U}_i$ factors. The corresponding CTVIR can be written as:

$$\hat{\mathbf{h}}_e(t,t_2)=\sum_{i=0}^{2^M-1}\delta\left(t_2-\frac{\Delta\tau_i}{2}\right)\mathbf{H}_i(t)\exp\left(-j\omega_0\frac{\Delta\tau_i}{2}\right), \quad (25)$$

where:

$$\mathbf{H}_i(t)=\prod_{m=0}^{M-1}\mathbf{U}^f_{M-m,d(M-m,i)}\mathbf{S}^s_{M-m}(t), \quad (26)$$

and $\Delta\tau_i$ are linear combinations of DGDs of all fibre segments calculated according to (15).

IV. Potential Applications

A. Relation between temporal distortions of received signal in a direct detection system and temporal behaviour of CTVIR

One potential application for the CTVIR of a fibre line with polarization scramblers can be analysis of efects of scrambling in direct detection systems. In such systems the received signal *r(t)* is proportional to the envelope of the optical signal (momentary optical power). This in turn is related to the complex envelope $\hat{\mathbf{w}}(t)$ by:

$$r(t)=\left|\hat{\mathbf{w}}_x(t)\right|^2+\left|\hat{\mathbf{w}}_y(t)\right|^2. \quad (27)$$

From (25) it can be easily stated that:

$$\hat{\mathbf{w}}(t)=\sum_{i=0}^{2^M-1}\mathbf{H}_i(t)\exp(-j0.5\omega_0\Delta\tau_i)\hat{\mathbf{v}}(t-0.5\Delta\tau_i). \quad (28)$$

According to (28) each output polarization mode $w_x(t)$ or $w_y(t)$ is an individual sum of 2^M-1 delayed and amplitude modulated (due to scrambling) copies of input polarization mode $v_x(t)$ and 2^M-1 delayed and amplitude modulated copies of input polarization mode $v_y(t)$:

$$\begin{aligned} y_x(t)&=\mathrm{Re}\left\{\sum_{i=0}^{2^M-1}g^i_{xx}(t)v_x\left(t-\frac{\Delta\tau_i}{2}\right)+\sum_{i=0}^{2^M-1}g^i_{xy}(t)v_y\left(t-\frac{\Delta\tau_i}{2}\right)\right\} \\ y_y(t)&=\mathrm{Re}\left\{\sum_{i=0}^{2^M-1}g^i_{yx}(t)v_x\left(t-\frac{\Delta\tau_i}{2}\right)+\sum_{i=0}^{2^M-1}g^i_{yy}(t)v_y\left(t-\frac{\Delta\tau_i}{2}\right)\right\} \end{aligned}, \quad (29)$$

where:

$$\begin{bmatrix} g^{(i)}_{xx}(t) & g^{(i)}_{xy}(t) \\ g^{(i)}_{yx}(t) & g^{(i)}_{yy}(t) \end{bmatrix}=\mathbf{H}_i(t)\exp(-j0.5\omega_0\Delta\tau_i). \quad (30)$$

The (29) shows that the fibre optic line with polarization scramblers resembles a system of 4 finite impulse response (FIR) filters with time varying coefficients. In each filter of the system there are at maximum 2^M delay sections with delays being linear combinations of DGDs of the fibres. Temporal variations of the coeffcients $g^i_{xx}(t)$, $g^i_{xy}(t)$, $g^i_{yx}(t)$, $g^i_{yy}(t)$ can result in time varying "allocation" of input signal to filter sections with different delays. Higher coefficients at sections with large delays than at those, with low delays, means that the energy of this specific output polarization mode propagates at lower speed. More delay difference between the two polarization modes that occurs at a given time results in increased distortion of the received signal *r(t)*.

Analysis of the $g^i_{xx}(t)$, $g^i_{xy}(t)$, $g^i_{yx}(t)$, $g^i_{yy}(t)$ could be the source of the knowledge on how polarization scrambling distorts received signal in a direct detection system. The method would possibly show the advantage of taking into account also PDL effect on scrambled signals.

However, (29) cannot be directly used for such an analysis because the $g^i_{xx}(t)x_x\left(t-\frac{\Delta\tau_i}{2}\right)$ components are not orthogonal so, the values of $g^i_{xx}(t)$, $g^i_{xy}(t)$, $g^i_{yx}(t)$, $g^i_{yy}(t)$ coefficients do not reflect the energy distribution in the domain of time delay.

B. Ortogonalization

One way of achieving othogonalization of the coeffcients $g^i_{xx}(t)$, $g^i_{xy}(t)$, $g^i_{yx}(t)$, $g^i_{yy}(t)$ is to convert (29) into a form that corresponds to FIR filters with equally spaced delays. Considering for a moment that $\hat{\mathbf{v}}(t)$ is sampled with the Nyquist interval *T* which is equal to the delay spacing, one can acknowledge the effect that each section of a FIR will filter an orthogonal sample.

A FIR filter is searched for that allows to calculate samples of the output signal through the relation:

$$\hat{\mathbf{w}}(kT, lT) = \sum_{p=-p\max}^{p\max} \hat{\mathbf{z}}(kT, pT)\hat{\mathbf{v}}((l-p)T), \qquad (31)$$

where p_{max} limits the number of coefficients in the FIR, *kT* and *lT* correspond to the running time and the "lag", respectively. The coeffcients of the $\hat{\mathbf{z}}(kT, pT)$ can be calculated when right sides of (28) and (31) are compared. This yields the following formula for the time varying coefficients:

$$\hat{\mathbf{z}}(kT, pT) = \sum_{i=0}^{2^M-1} \mathbf{H}_i(kT) e^{-j\omega_0 \frac{\Delta\tau_i}{2}} Sa(\pi f(pT - 0.5\Delta\tau_i)), \qquad (32)$$

where: *f* is the sampling frequency.

C. Experiment

The temporal evolution of time varying coeffcients of a filter $\hat{\mathbf{z}}(kT, pT)$ representing the CTVIR of a fibre line with polarization scramblers were observed in a simulated test experiment. A 5 segment fibre line was simulated, each segment with equal DGD value of 7.5ps. This corresponded to the maximum compound DGD 37.5ps which was 37.5% of the bit signalling interval (100 ps). 0.5dB PDL was assumed for each fibre segment. Also every fibre segment was preceded by a 3-axis polarization scrambler driven by sinusoidal signals. The frequencies of the signals driving polarization rotations were set to 50, 51 and 52 MHz – in the first scrambler, 53, 54 and 55 in the second one, and so on with the last scrambler driven by 62, 63 and 64 MHz . The modulation format was RZ with 33% duty cycle at 10Gbps data transmission speed. The sequence consisted only of ones (…11111…) for better visual inspection of the results. The signal received in a direct detection system is assumed proportional to the momentary total power of both polarization components. In the orthogonal representation of the CTVIR the delay spacing in the FIR was selected 12.5ps. The number of sections was 12 per filter.

The results are illustrated in the Fig. 1 and 2. In order to reduce the amount of presented data $|g^i_{xx}(t) + g^i_{xy}(t)|$ and $|g^i_{yx}(t) + g^i_{yy}(t)|$ were chosen as indicators of the $\hat{\mathbf{z}}(kT, pT)$ filter behaviour. This corresponds to the amplitudes (envelopes) of both output polarizations provided input polarizations are excited equally.

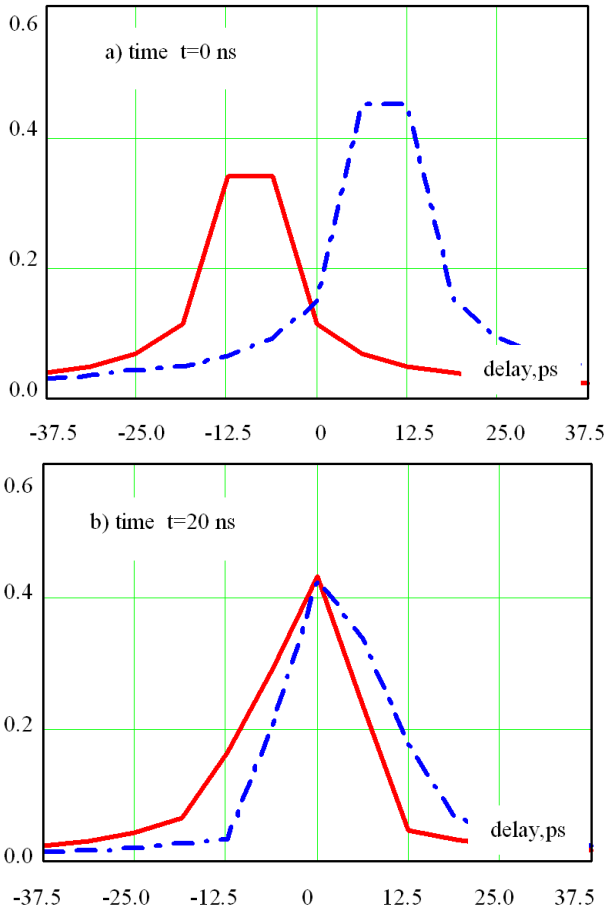


Fig 1. Magnitudes of coefficients of the FIR filters which model the CTVIR of the simulated fibre optic line with polarization scramblers.

In the Fig 1 a-b) $|g^i_{xx}(t) + g^i_{xy}(t)$ (red line) and $|g^i_{yx}(t) + g^i_{yy}(t)|$ (blue line) are shown as functions of the index of a FIR filter section. However, instead of using index numbers the horizontal axes are scaled in time units representing the time delay that are introduced by particular sections. The vertical axes represent values of the magnitudes. The a) and b) versions correspond to distinct time instances at which the FIR filters are inspected.

In the Fig. 2 the contour plots a-b) show temporal evolutions of $|g^i_{xx}(t) + g^i_{xy}(t)|$ and $|g^i_{yx}(t) + g^i_{yy}(t)|$ due to scrambling. The horizontal axes represent the running time in nanoseconds. The vertical axes indicate the delay in picoseconds, associated with a particular section of a filter. For the comparison the plot of the momentary optical power at the receiver end is given in the Fig. 2 c).

The Fig. 1 and Fig. 2 c) seem confirming the intuitive idea on existence of a functional connection between distortions of the enveope and the spread of distributions of the coefficients that describe transfer of input polarization components to the output polarization components. This functional connection can also be deduced from the Fig 2 a-b).

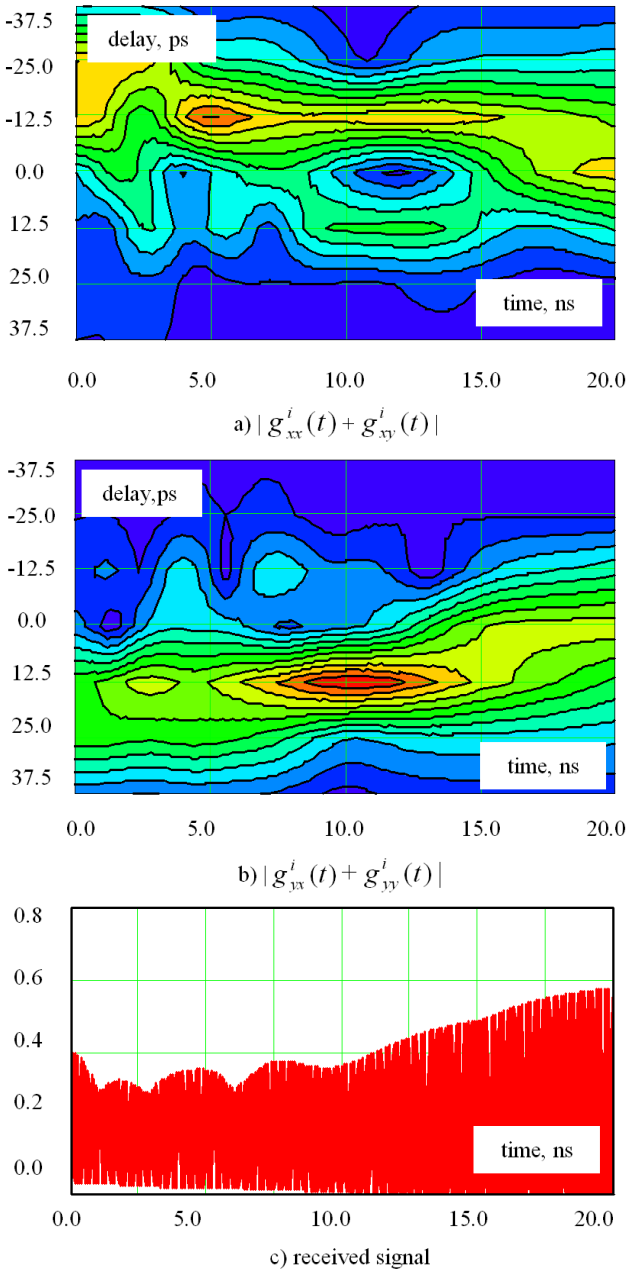


Fig 2. a-b) Temporal evolution of the filter representing CTVIR of the simulated fibre optic line with polarization scramblers; c) temporal evolution of the envelope of optical signal.

For example in the 0..5 ns time interval the delays of the two polarizations differ significantly and the result manifests in the Fig. 2 c) as low amplidude received pulses (the envelope of the signal undergoes significant deterioration). The opposit effect which can be explained by the same mechanism can be noticed in the last 5 nanoseconds in the above example when the filters transfer polarization components with low or even zero delays. The resultant momentary optical power is relatively high (low deterioration of the envelope).

For the remaining time interval it is not so easy to find the expected functional relations. In fact, the undistorted signal can be observed not only when the both polarizations are undelayed but also when majority of the energy is transferred to a single polarization. This may suggest that the relation of interest is not strightforward.

This preliminary analysis only highlights the possibilities of using the CTVIR to analyse the influence of scrambling on momentary optical power at the end of a fibre optic line in a direct detection system. The exact relations of interest as well as their practical applicability shall remain subjects of further studies.

V. CONCLUSIONS

A complex time-variant impulse response - CTVIR can be considered a tool for analysis of a temporal behaviour of a fibre optic transmission line with polarization scramblers driven by real world signals. The CTVIR can be modelled by a set of 4 finite impulse response filters with time varying coeffcients. It seems that there can be developed a model of optical signal momentary power distortions due to scrambling that will exploit the knowledge on temporal evolution of these coeffcients. The possible functional relations and their practical applicability need further studies.

REFERENCES

[1] R. Leppla, "Outage probability of high speed transmission systems: The impact of PMD in comparison to other disruptive effects," Optical Fiber Communication Conference, 2001 OSA Technical Digest Series (Optical Society of America, 2001), paper WDD38.

[2] H. Bülow, C. Xie,A. Klekamp, X. Liu, B. Franz, "PMD Compensation/Mitigation Techniques for High-Speed Optical Transport", Bell Labs Technical Journal 14(1), 2009, pp. 105–124.

[3] P. K. Kondamuri, C. Allen, and D. L. Richards, "Variation of PMD-Induced Outage Rates and Durations with Link Length on Buried Standard Single-Mode Fibers," Optical Fiber Communication Conference and Exposition and The National Fiber Optic Engineers Conference, Technical Digest (CD), paper OThX3.

[4] J. Li, G. Biondini, H. Kogelnik, P. J. Winzer, "Noncompliant Capacity Ratio for Systems With an Arbitrary Number of Polarization Hinges", Journal of Lightwave Technology, Vol. 26, No. 14, July 15, 2008, pp. 2110-2117.

[5] R.-J. Essiambre, G. Kramer, P. J. Winzer, G. J. Foschini, B. Goebel, „Capacity Limits of Optical Fiber Networks", Journal of Lightwave Technology, Vol. 28, Issue 4, 2010, pp. 662-701.

[6] J. P. Gordon, H. Kogelnik, „PMD fundamentals: Polarization mode dispersion in optical fibers", Proceedings of the National Academy of Sciences of the United States of America, Vol. 97, No. 9, 2000, pp. 4541-50.

[7] W.A. Ainsworth, "Time variant filters and analytic signals", IEEE Transactions on Inf. Theory, Vol 13, Issue 2, 1967, pp. 331-333

Upgrading the Optical Access Network Architectures

Peter Kubizniak
Institute of Telecommunications
Faculty of Electrical Engineering and Information Technology, Slovak University of Technology
Bratislava, Slovakia
e-mail: kubizniak@ut.fei.stuba.sk

Abstract — **The article provides a basic insight into optical access network design issue. Beside other access solution, the focus is on passive optical networks, namely TDM PONs and WDM PONs. Several architectures are listed and their improvement proposed based on examining key factors like economic impact, system range, scalability etc.**

Keywords - optical network, architecture, TDM PON, WDM PON, GPON, EPON, NGA

I. INTRODUCTION

Optical networks are about to overtake the dominant position of metal networks in access networks. Fiber-to-the-home (FTTH) is seen as the ultimate and most futureproof access solution. In consequence, this means building a completely new access network, thus requiring enormous investment and potentially allowing new business models. In the long run this will enable next-generation optical access (NGOA) networks where the access network is not limited to the first mile but will potentially further extend beyond today's central offices (COs).

At the begining, the main options for accessing the network via optical fiber are given. The next section briefly addresses several solution for improving the most-spreaded time division mupliplexed (TDM) passive optical networks (PON). The latter part provides survey of selected wavelength division multiplexed (WDM) PON architectures. The conclusion looks up to the future work.

II. OPTIONS FOR ACCESSING OPTICAL NETWORKS

Optical access networks can be deployed using different architectures as shown in Fig. 1. In a point-to-point (PtP) architecture all subscribers are connected to an access node (e.g., CO – Central Office) via dedicated fibers. Today's PtP deployments are mainly based on Ethernet technology using Ethernet switches in the access node with a high port density. The network termination at the subscriber site is realized with media converters (e.g., 100Base-TX, 100Base-BX) or mini-switches. The PtP architecture with its dedicated fiber connections requires a very high number of fibers in the whole access network, which causes high costs for fiber rollout and handling. In addition, each connection requires two interfaces, which cause high footprint and power consumption. In order to reduce the high number of fibers in the access network, point-to-multipoint (PtMP) architectures can be used. PtMP architecture offers one or more additional aggregation layers between the subscriber location and CO. In general, two PtMP architectures can be distinguished: the active optical network (AON) and PON.

The AON is determined by an active aggregation element (e.g., Ethernet switch) in the first mile. In one AON concept an Ethernet switch is located at the street cabinet (2a), whereas in the second variant an Ethernet switch is used at the building location (2b). On one hand the AON allows a reduction of the fiber count in the access network compared to a PtP solution, but on the other hand it is not able to decrease the number of required interfaces, so it is virtually impossible to reduce the footprint and power consumption.

In contrast to an AON, PON aggregation is based on passive components such as optical power splitters or WDM (de)multiplexers. Today, typically 32–64 optical network units/terminals (ONUs/ONTs) are connected to one PON port at the optical line termination (OLT). This means that PON architecture enables fiber reduction as well as optimization of the footprint and power consumption compared to PtP architecture. [1]

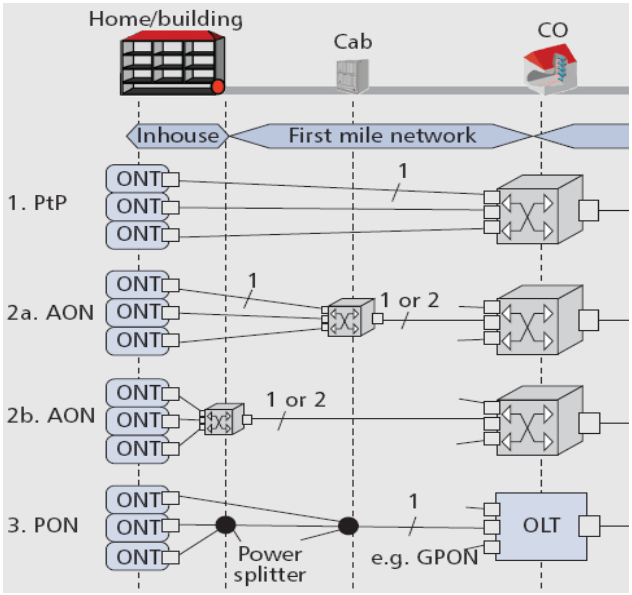


Figure 1. Optical access networks architectures

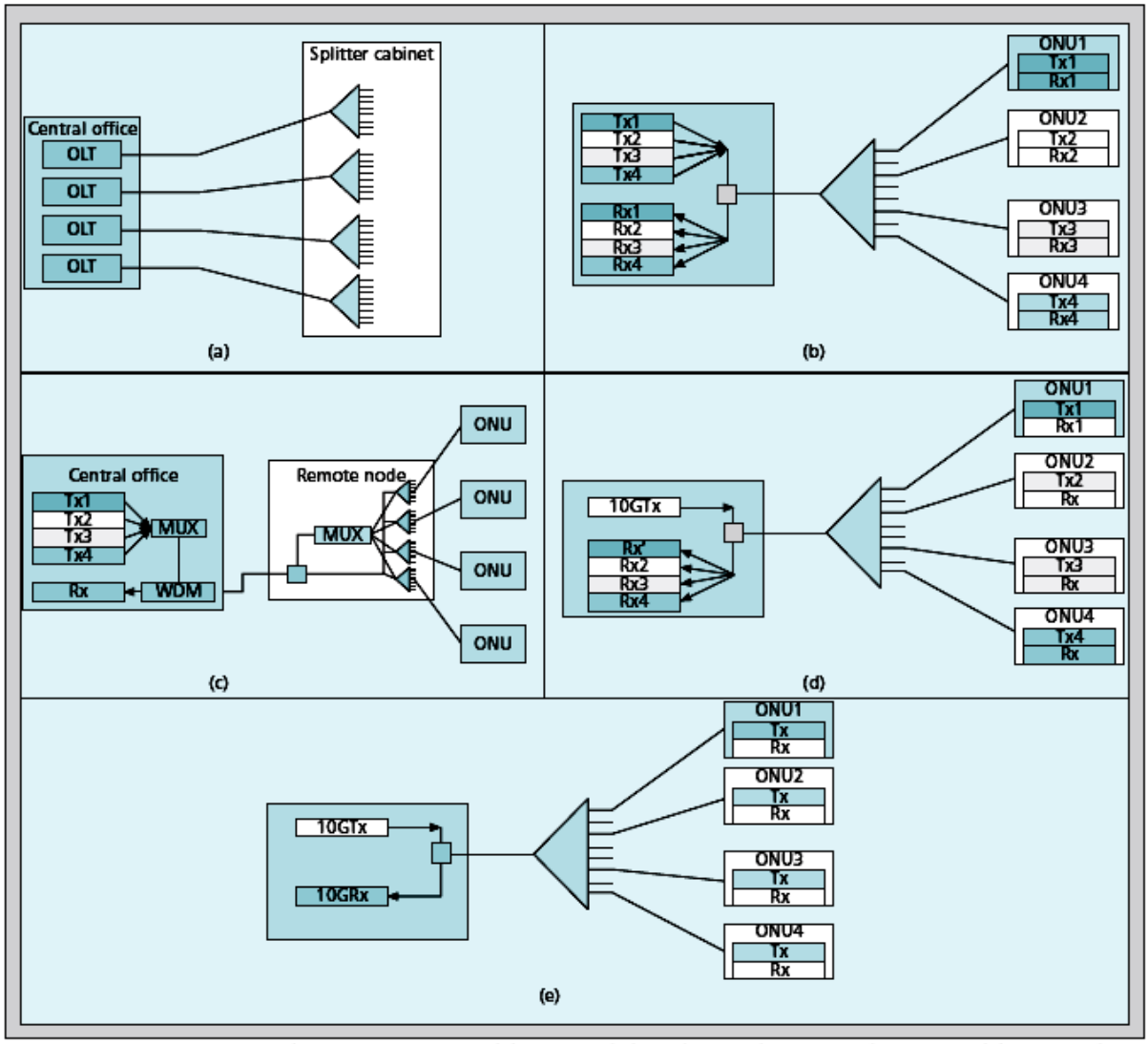


Figure 2. NGA1 architectures: *a) GPON with bidirectional physical PON split reduction; b) GPON with bidirectional wavelength PON split reduction; c) GPON with downstream wavelength PON split reduction; d) XG-PON1 with 10G downstream, Nx2.5G upstream; e) XG-PON2 with 10G symmetric.*

III. UPGRADING TDM PON ARCHITECTURES

A. Upgrading TDM PON (GPON) to NGA

The Full Service Access Network (FSAN) is now studying the next-generation access (NGA). The objective of NGA is to facilitate high bandwidth provision, large split ratio, and extended network reach. FSAN has planned two stages of NGA evolution: NGA1 and NGA2. NGA1 focuses on PON technologies that are compatible with GPON standards (ITU-T G.984 series) and compatible with the current optical distribution network (ODN) as well. In contrast, NGA2 is a long-term solution with an entirely new optical network type. The objective of NGA2 is to provision an independent PON scheme, without being constrained by the GPON standards and the currently deployed outside plant.

This section focuses on candidate NGA1 network architectures. Five typical candidate network architectures have been proposed for NGA1. They employ different methods in achieving the increase of the upstream and downstream bandwidth per ONU. On the one hand, given that NGA1 is compatible with the GPON standards and the currently deployed ODN, candidate NGA1 architectures try to leave the ODN part untouched but upgrade its devices at head ends including OLTs and ONUs. On the other hand, like GPON, NGA1 still employs the TDM principle in multiplexing and demultiplexing traffic onto a wavelength channel. [2]

Among the five candidate architectures, NGA1-4 is a promising and economical architecture to meet the future bandwidth requirement in NGA1. First, the upstream and downstream bandwidths are increased to 2.5 Gb/s and 10 Gb/s, respectively. The increased bandwidths are potentially able to accommodate the future bandwidth-consuming applications in NGA1. Second, the 2.5-Gb/s burst mode receiver requires lower cost compared to the 10-Gb/s burst mode receiver in NGA1-5, making NGA1- 4 a more economical solution for a GPON upgrade. Essentially, the upgraded systems are still TDM systems. By abstraction, the typical 32 ONUs in GPON

are divided into multiple virtual groups, where the ONUs in each group share a link in TDM fashion. Hence, the five candidate architectures can be regarded as TDM systems with different link rates and different numbers of shared ONUs. In this article, we assume that ONUs are evenly divided among the upstream or downstream channels. The upstream and downstream link rates of GPON are 1.244 Gb/s and 2.488 Gb/s, respectively.

Table 1 summarizes the characteristics of the abstracted TDM systems in both directions for the five architectures. For the upstream, NGA1-1 and NGA1-2 are abstracted as four virtual groups, each of which has eight ONUs sharing a 1.244-Gb/s link; NGA1-3 is abstracted as one virtual group with 32 ONUs sharing a 1.244-Gb/s link; NGA1-4 is abstracted as one virtual group with 32 ONUs sharing a 2.488 Gb/s-link if one 2.488 Gb/s link is used, and two virtual groups each of which with 16 ONUs sharing a 2.488-Gb/s link if two 2.488- Gb/s links are used; NGA1-5 is abstracted as one virtual group with 32 ONUs sharing a 10- Gb/s link. For the downstream, ONUs in NGA1-1, NGA1-2, and NGA1-3 are abstracted as four virtual groups, each of which is with eight ONUs sharing a 2.488-Gb/s link; ONUs NGA1-4 and NGA1-5 are abstracted as one virtual group with 32 ONUs sharing a 10-Gb/s link. [2]

TABLE I. ABSTRACTION OF NGA1

	NGA1-	***1***	***2***	***3***	***4***	***5***
Upstream	**Virtual groups**	4	4	1	1(2)	1
	ONUs/group	8	8	32	32(16)	32
	Rate [Gb/s]	1.224	1.224	1.224	2.488	10
Downstream	**Virtual groups**	4	4	4	1	1
	ONUs/group	8	8	8	32	32
	Rate [Gb/s]	2.488	2.488	2.488	10	10

B. Upgrading TDM PON (EPON) to WDM

At present, EPONs are single-channel systems where ONUs share the bandwidth of the upstream channel by means of Time Division Multiplexing (TDM). As the number of subscribers as well as the traffic volume per subscriber increase, EPONs that use a single wavelength in each fiber will very likely be unable to cope with the growing traffic demands of the ONUs. To use the fiber infrastructure of PONs in a more cost-effective fashion, research has recently begun to revisit multichannel PONs that deploy Wavelength Division Multiplexing (WDM). A WDM architecture for an EPON would be as follows:

- The OLT consists of an array of fixed laser/receivers.
- The ONUs consists of either an array of fixed laser/receivers or one or more tunable laser/receivers.

All of the above will most likely be upgraded incrementally over long periods of time. In service provider networks managing ONUs with different WDM capabilities can be extremely difficult. This is due to the fact that service providers will have to record the types of ONUs deployed as well as maintain spare parts inventories for each of the laser/receiver components. Given this, it is more likely that service providers will do the following:

- Use either tunable laser/receivers or fixed laser/receiver arrays, not both.
- Upgrade in only a discrete number of incremental stages. (This eases ONU management.)

Regardless of the above, the evolutionary upgrade to WDM on EPONs should not impose any particular WDM architecture. Thus allowing these decisions to be dictated by economics and service provider preferences. Anyway, the aim is to achieve the following goals:

- Support the addition of WDM capabilities in an EPON on an evolutionary basis
- Not impose a particular WDM architecture on the ONUs
- Enable the OLT to schedule transmission to and reception from ONUs on any wavelength supported by that ONU. [2]

IV. WDM PON ARCHITECTURES

Several WDM-PON architectures have been proposed for providing scalability that is lacking in traditional PONs. Straightforward approach to increase the capacity of a TDM-PON is to assign a separate wavelength channel to each user. However, this solution provides high performance at a high cost; e.g., whenever a user shuts down his connection, then the corresponding transceiver at the OLT is idle, and will not be able to support other users, in which case the network resource is left unused. In this section, we review some of the representative WDM-PON architectures which have been proposed in the literature. [3]

A. Composite PON (CPON)

To solve the scalability limitation of traditional PONs, one of the earliest WDM-PON architecture proposals based on the arrayed waveguide grating (AWG) concept employed WDM in the 1550–nm band in the downstream direction and a single upstream wavelength in the 1300–nm band shared through time-division multiple access (TDMA).While an earlier design of this architecture used separate fibers for upstream and downstream, an integrated type of device was proposed that performed WDM routing in one wavelength window and wavelength independent power combining in a second window on a single fiber through CWDM. This architecture has been referred to as the composite PON (CPON). CPON is limited in the fact that a singlefrequency laser, such as a distributed-feedback (DFB) laser diode (LD) at the ONU, may be economically prohibitive. [3]

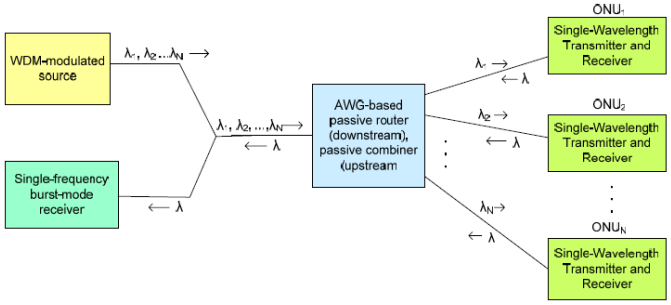


Figure 3. Architecture of CPON

B. *Local Access Router Network (LARNET)*

The LARNET (local access router network) architecture attempts to work around the limitation in CPON by employing a broad-spectrum source at the ONU, such as an inexpensive light-emitting diode (LED) whose spectrum is sliced by the AWG-based router into different optical bands in the upstream direction. The edge-emitting LED emits a broad spectrum of wavelengths centered around a single wavelength, as compared to the DFB laser, which emits only one wavelength of light. An advantage of using an edge-emitting LED is that each ONU can have the same device, as opposed to employing DFB lasers at each ONU, which may require for the DFBs to be different.

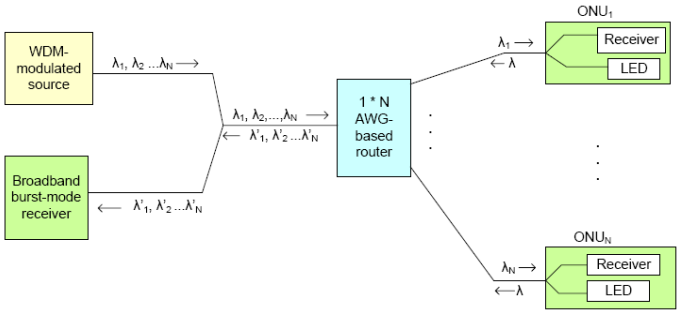


Figure 4. Architecture of LARNET.

Edge-emitting LEDs are much cheaper compared to DFB lasers, so they reduce the cost of the ONU. The limitation with this architecture is that spectrally slicing a broad-spectrum source by an AWG may lead to high power loss. Therefore, the distance from the OLT to the ONU is considerably reduced in LARNET. [3]

C. *Remote Interrogation of Terminal Network (RITENET)*

In RITENET (remote interrogation of terminal network), the light at the ONU is split by a passive tap with a portion of the light detected by the receiver. The remainder is looped back toward the CO through a modulator. The signal from the OLT is shared for downstream and upstream through time sharing. A 2×2N AWG-based router is used to route the wavelengths.

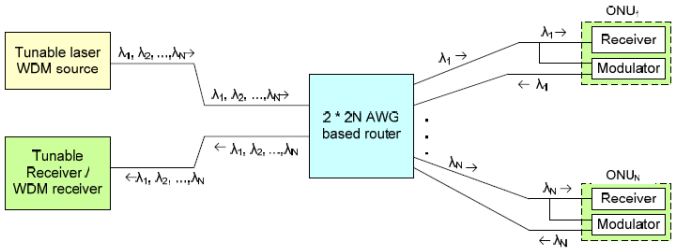


Figure 5. Architecture of RITENET

Since the same optical channel is used for both upstream and downstream, they must be separated on two different fibers. Figure 5 shows the architecture of RITENET. The number of fibers employed in the architecture is doubled, thus doubling the cost of deployment and maintenance. Also employing either a tunable laser or an array of transmitters and receivers at the OLT makes RITENET a more expensive architecture compared to CPON and LARNET. The architecture, however, has some advantages, such as the availability of symmetrical bandwidth in the downstream and upstream directions. Also, the upstream signal does not suffer the spectral-slicing loss seen in LARNET. [3]

D. *Multistage AWG-Based WDM PON Architecture*

The above architectures have two main limitations: (1) difficulty in scaling the number of ONUs once the network is laid out, and (2) limited number of users, because the fabrication technology limits the AWG size. To overcome these limitations, a multistage WDM-PON exploits the periodic routing property of the AWG so that the reuse of a given wavelength for more than one subscriber is possible. This architecture provides scalability in bandwidth as well as in number of users, either by employing additional wavelengths at the CO or by cascading multiple stages of AWGs with increasing AWG coarseness at each stage. [3]

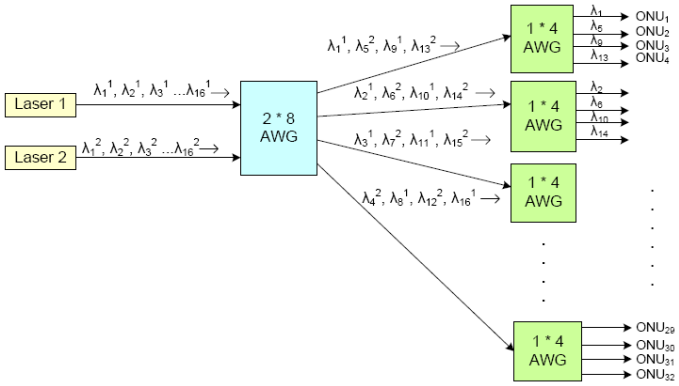


Figure 6. Architecture of Multistage AWG-based WDM PON

E. SUCCESS-DWA PON Architecture

The Stanford University access dynamic wavelength assignment (SUCCESS-DWA) PON architecture aims to offer scalability by employing dynamic wavelength allocation (DWA) to provide bandwidth sharing across multiple physical PONs. This is done while the existing arbitrary field-deployed PON remains intact. The network architecture employs tunable lasers (TLs) and the AWG at the CO, while WDM filters and a burst-mode receiver are employed within the ONUs. The upstream and downstream traffic is separated by a wideband WDM filter residing between the AWG and the PON. This architecture requires that all wavelengths from the OLT can reach all ONUs across separate physical PONs.

The architecture provides scalability by initially deploying one TL and one AWG in the CO, which services multiple subscribers across several PONs. As the demand grows, the architecture can be scaled by adding more TLs to the AWG or by adding another AWG along with more TLs. [4]

F. The SUCCESS Hybrid WDM/TDM PON or SUCCESSHPON

The SUCCESS-HPON architecture proposes a migration path from TDM to WDM-PON using a centralized light sources (CLS) approach and tunable WDM component sharing for costefficiency. The overall architecture of SUCESS-HPON, including TDM-PONs and WDM-PONs as its subnetworks, is shown in Fig. 7. A single-fiber collector ring with stars attached to it formulates the basic topology. The collector ring strings up Remote Nodes (RNs), which are the centers of the stars. The ONUs attached to the RN on west side of the ring talk and listen to the transceiver on the west side of OLT, and likewise for the ONU attached to the RNs on the east side of the ring.

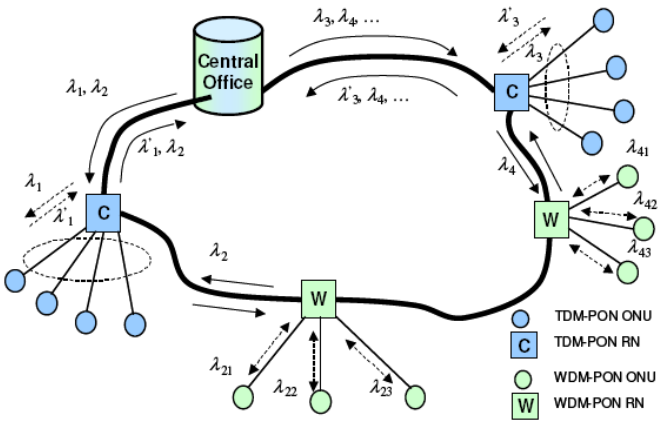


Figure 7. Architecture of SUCCESSHPON

There is a point-to-point WDM connection between the OLT and each RN. No wavelength is reused on the collector ring. Optional semi-passive RNs may be used to sense fiber cuts and flip the orientation, if extra reliability is desired. [5]

V. CONCLUSION

There are many aspects when building or upgrading the optical access network like the costs, scalability, and maintenance that need to be considered. The aim of this paper was to step inside the issue of designing the network architecture tailored to specific requirements. This should give the reader basic orientation which way to choose when looking for a solution for either building a new topology or improving the existing one.

In our work, we analyze the possibilities for effective utilization of optical fiber in optical access networks. The main approach to achieve efficiency in utilizing the physical capability of optical fiber is to implement suitable mechanism for bandwidth allocation. We proposed basic algorithms for bandwidth allocation in [6] used mainly in TDM PONs and algorithms for wavelengths assignment in [7] used mainly in WDM PONs.

Analyzing these architectures will help us to choose the architecture for our next research, which will be applying and examining these algorithms.

ACKNOWLEDGMENT

This work is a part of research activities conducted at Slovak University of Technology Bratislava, Faculty of Electrical Engineering and Information Technology, Department of Telecommunications, within the scope of the project VEGA No. 1/0106/11 “Analysis and proposal for advanced optical access networks in the NGN converged infrastructure utilizing fixed transmission media for supporting multimedia services”.

REFERENCES

[1] D. Breuer, F. Geilhardt, R. Hulsermann, M. Kind, and C. Lang, “Opportunities for Next-Generation Optical Access,” Deutsche Telekom Laboratiories, IEEE Communications Magazine, pp. S16-S24, February 2011.

[2] J. Zhang, and N. Ansari, “Next-Generation PONs: A Performance Investigation of Candidate Architectures for Next-Generation Access Stage 1,” New Jersey Institute of Technology, IEEE Communications Magazine, pp. 49-57, August 2009.

[3] A. Banerjee, Y. Park, F. Clarke, H. Song, and G. Kramer, “Wavelength-division-muliplexed passive optical networks (WDM-PON) technologies for broadband access: a review,” Optical Society of America, Journal of Optical Networking, vol.4, pp. 737-758, November 2005

[4] Y. Hsueh, M. S. Rogge, W. Shaw, and L.G. Kazovsky,”SUCCESS-DWA: A highly scalable and cost-effective optical access network,” Standford University, IEEE Communications Magazine, pp. S24-S30, August 2004.

[5] D. Gutierrez, W. Shaw, F. An, Y. Hsueh, and M. Rogge,”Next Generation Optical Access Networks,” Standford University [Invited Paper]

[6] P. Kubizniak,”The bandwidth utilization in the passive optical networks,” Conference proceedings – IMEA 2010, Seč u Chrudimi, Czech Republic

[7] P. Kubizniak,”Effective utilization of wavelengths in passive optical networks,” Conference proceedings - ELEKTRO 2010, Žilina, Slovakia

Fog attenuation influence on Bit Error Ratio of Free Space Optics link

Jan Vitasek, Jan Latal, Stanislav Hejduk, Jiri Bocheza, Petr Koudelka, Petr Siska, Jan Skapa,Vladimir Vasinek
Department of Telecommunications
VŠB – Technical University Ostrava
Ostrava, Czech Republic
jan.vitasek@vsb.cz

***Abstract*—The fog is an atmospherical phenomenon which influences the FSO communication the most. In this article is described how the fog affects bit error ratio of FSO link and how the eye diagrams look like. It was set a FSO link in software OptiSystem. This software enables to set different parameters of link and measure and display different values. For this simulation it was used BER measurement and Eye diagram.**

Keywords-fog; attenuation coefficient; Q-factor; BER; eye diagram

I. INTRODUCTION

The fog is water that has condensed close to ground level, producing a cloud of very small droplets that reduce visibility [1]. Fog consists of small water droplets or tiny ice crystals suspended in the air. These small droplets cause dispersion of laser beam therefore all transmitted power can not be received by receiver. This effect is more significant by fog than by rain because the fog droplet is considerably smaller than rain droplet. A diameter of fog droplet can be from 1 to 20 μm.

II. FOG ATTENUATION

The fog significant influences the atmosphere transmittance and therefore the attenuation of free space optics (FSO) link. The atmosphere transmittance is described by the help of Beer Lambert law.

$$T(\lambda,l)=\frac{P(\lambda,l)}{P(\lambda,0)}=e^{-[\gamma(\lambda)l]}, \quad (1)$$

where $T(\lambda)$ is atmosphere transmittance for wavelength λ, $P(\lambda, l)$ means a power level of optical signal at the distance l from the transmitter, $P(\lambda, 0)$ is a source optical power and $\gamma(\lambda)$ is an extinction coefficient. From extinction coefficient $\gamma(\lambda)$ is easily possible to obtain an attenuation coefficient. The resulting equation to calculate the attenuation coefficient is:

$$\alpha_{fog}=\frac{17}{V_m\cdot\left(\frac{\lambda_r}{\lambda_0}\right)^q}, \quad (2)$$

where α_{fog} is an attenuation of fog, unit is dB/km [2]. V_m is a meteorological visibility, unit is km, λ_r is reference wavelength (555 nm) and λ_0 is a wavelength of laser beam, unit is nm. Parameter q is described by several models which were set according empirical measurement.

The best known are Kruse model a Kim model. According Kruse model the parameter q is set by next values [3].

TABLE I. Q VALUES OF KRUSE MODEL

Table of q values – Kruse model	
V_m [km]	q
$V_m > 50$ km	1.6
6 km $< V_m \leq 50$ km	1.3
$V_m \leq 6$ km	$0.585\ V_m^{1/3}$

Kim model sets next values of parameter q [3].

TABLE II. Q VALUES OF KIM MODEL

Table of q values – Kim model	
V_m [km]	q
$V_m > 50$ km	1.6
6 km $< V_m \leq 50$ km	1.3
1 km $< V_m \leq 6$ km	$0.16\ V_m + 0.34$
0.5 km $< V_m \leq 1$ km	$V_m - 0.5$
$V_m \leq 0.5$ km	0

III. Q-FACTOR, EYE DIAGRAM AND BIT ERROR RATIO (BER)

A. Q-factor

Q-factor represents a quality of digital signal with a view to signal-noise ratio. It contains all physical deteriorations which work upon signal like is noise, nonlinear effects and so on. These deteriorations degrade the signal and make bit errors. Therefore the greater Q-factor value means greater signal-noise ratio and then less bit error ratio. The equation (3) describes Q-factor calculation of optical signal:

$$Q = \frac{I_1 - I_0}{\sigma_1 + \sigma_0}, \tag{3}$$

where I_1 is logical level „1“, I_0 is logical level „0“, σ_1 is standard deviation of logical level „1“ a σ_0 is standard deviation of logical level „0“ [5].

B. Bit error ratio

In real digital communication systems there is set a decision level in which time segment it should sample and whether the sampled value represents logical level „1“ or logical level „0“. This decision is affected by noise and distortion, there is nonzero probability of erroneous decision. Therefore the quality of received signal relates to bit error ratio, which is the main quality parameter of the system [5]. The equation (4) shows the relationship between Q-factor and bit error ratio BER.

$$BER = \frac{1}{2}\operatorname{erfc}\left(\frac{Q}{\sqrt{2}}\right) \approx \frac{\exp\left(\frac{-Q^2}{2}\right)}{Q\sqrt{2\pi}}. \tag{4}$$

C. Eye diagram

The eye diagram represents the superposition of the all mutually overlapped bits in signal. The eye diagram shows us two types of undesirable effects. The one effect is an intersymbol interference (ISI) and the other is jitter. ISI is caused by overlapping of modulation impulses. The jitter is defined as short-term deviation of digital signal from its ideal position. The wider eye opening means smaller attenuation or distortion and then higher quality signal [5].

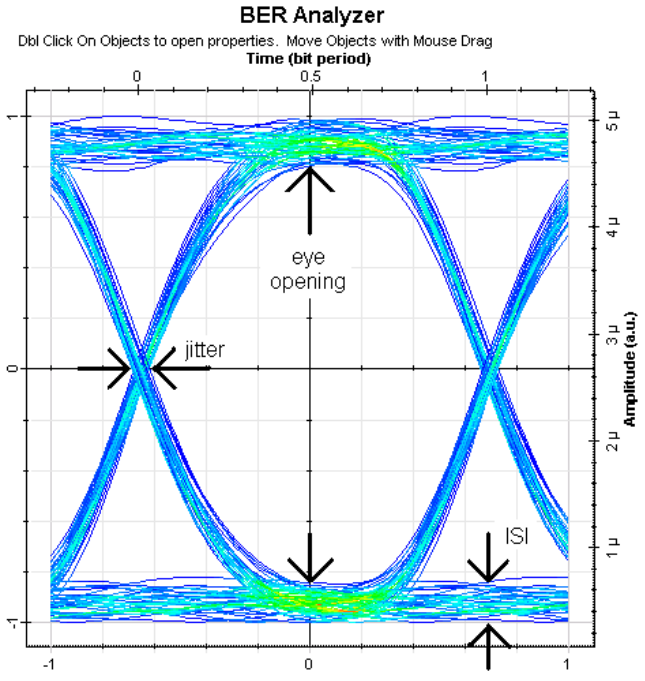


Figure 1. Eye diagram

IV. Simulation in OptiSystem aplication

OptiSystem is an innovative optical communication system simulation package that designs, tests, and optimizes virtually any type of optical link from the physical layer to the transport layer according OSI model.

OptiSystem is a stand-alone product that does not rely on other simulation frameworks. It is a system level simulator based on the realistic modeling of optic communication systems. It possesses a powerful new simulation environment and a truly hierarchical definition of components and systems. Its capabilities can be extended easily with the addition of user components, and can be seamlessly interfaced to a wide range of tools [6].

In this program there was set a FSO link. At the start was CW laser with wavelength 850 nm and optical power 0dBm. Pseudo-random bit sequence generator produced data which were modulated on the optical carrier wave by using external modulator. It was used NRZ modulation. This signal was transferred by the atmosphere, the FSO link length was set 1 km. At the end was placed PIN photodetector and after it there was BER analyzer which measured Q-factor, BER and Eye diagram.

Other parameters of FSO link were subsequent. Transmitter aperture diameter was 5 cm, receiver aperture diameter 10 cm, beam divergence 0.25 mrad.

How is described above the fog causes atmosphere attenuation. The attenuation coefficient α_{fog} was calculated according equation (2). The parameter q is set by Tab. II., thereby the limiting attenuation coefficient values were set for given visibilities.

TABLE III. Attenuation coefficient according Kim model

Table of α_{fog} values – Kim model	
V_m [km]	α_{fog} [dB/km]
$V_m > 50$ km	<0.17
6 km < $V_m \leq 50$ km	0.17-1.61
1 km < $V_m \leq 6$ km	1.61-13.67
0.5 km < $V_m \leq 1$ km	13.67-34.00
$V_m \leq 0.5$ km	>34.00

At first it was investigated Q-factor and bit error ratio BER dependence on the attenuation coefficient α_{fog} variation. The all parameters of FSO link were set to above described values. The attenuation coefficient α_{fog} variation means fog density variation, the greater value, the denser fog.

Subsequently it was investigated Q-factor and bit error ratio BER dependence on the FSO link length variation by constant attenuation coefficient α_{fog} = 8 dB/km. The target is to know how the BER of received signal changes in dependence on increasing length.

V. SIMULATION RESULTS

Pseudo-random bit sequence generator generated the following NRZ data. These data were modulated on the optical carrier wave and transmitted by FSO link.

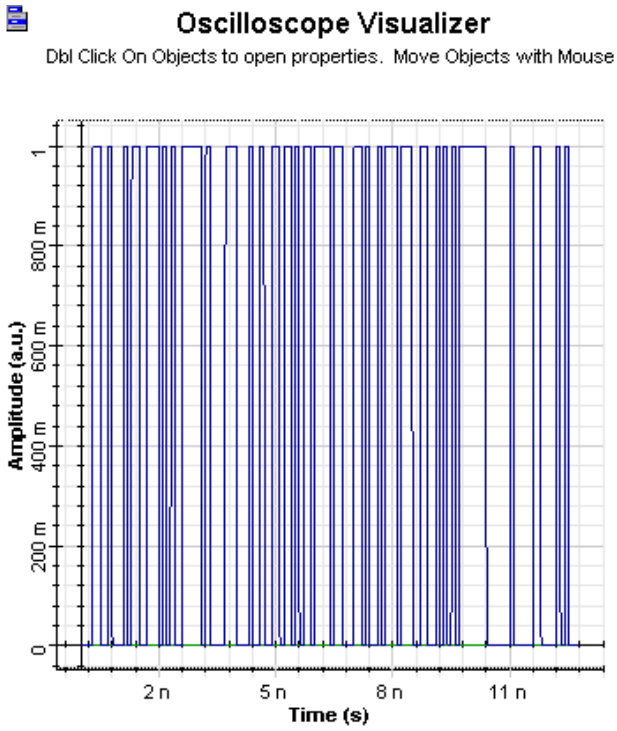


Figure 2. Pseudo-random NRZ data

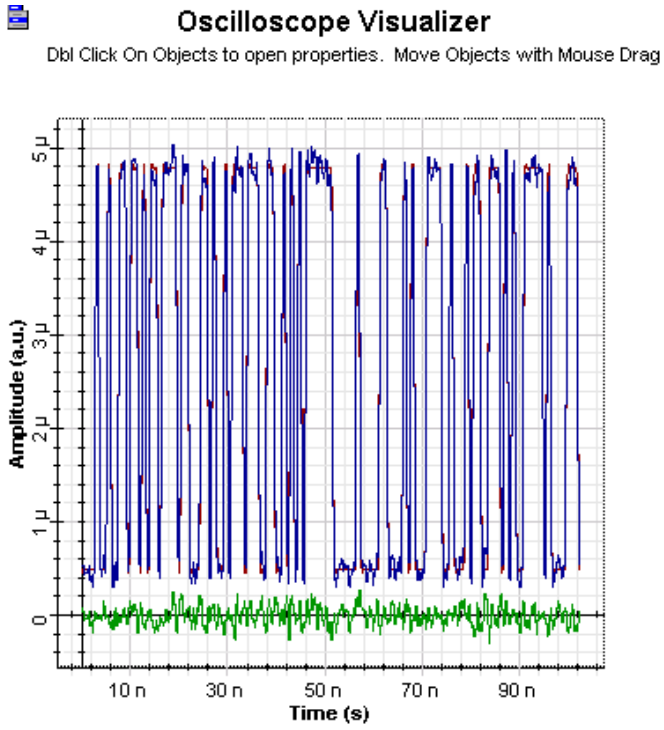


Figure 3. Received data

PIN fotodetector converts optical signal back to the data. PIN photodetector was set subsequently: responsivity 1 A/W, dark current 10 nA, thermal noise 10^{-26} W/Hz. These parameters influence obtained results minimally. Higher values of thermal noise make worse data reading but it does not influence FSO link. The Fig. 3 was obtained by setting FSO link length 1 km and attenuation coefficient α_{fog} = 13.67 dB/km. The green curve is noise, blue curve is data signal+noise, red curve is data signal.

According Tab. III. four limiting attenuation coefficients were chosen and for these coefficients the eye diagrams were displayed. These coefficients were α_{fog} = 0.17, 1.61, 13.67, 34.00 dB/km. FSO link length was 1 km.

Fig. 4 shows the eye diagram pro FSO link length 1 km and attenuation coefficient α_{fog} = 0.17 dB/km. This attenuation coefficient represents pure atmosphere [7]. We can see the eye opening is wide therefore the BER of FSO link is minimum. The same can be said about intersymbol interference ISI and jitter which are very small. The BER component in OptiSystem allows measure max. Q-factor, min. BER, Eye height, it is not able to measure ISI and jitter. Therefore ISI and jitter dependences on attenuation coefficient were not measured.

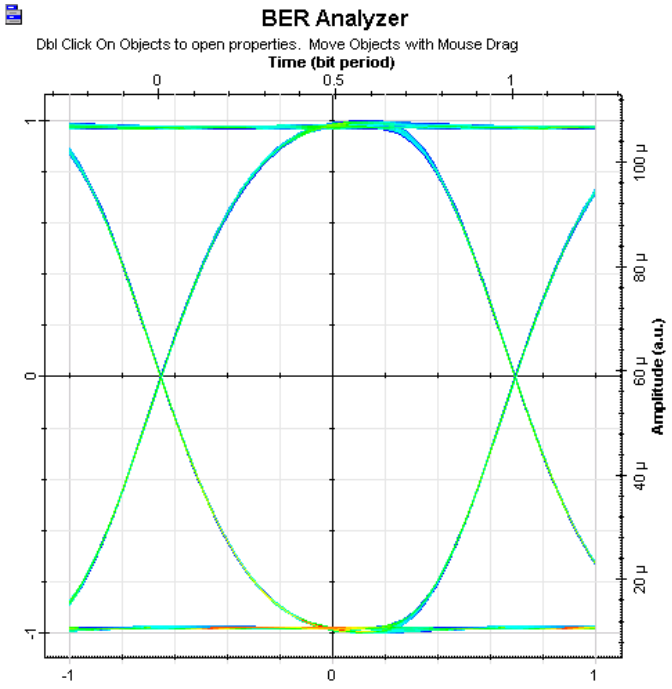


Figure 4. Eye diagram, α_{fog} = 0.17 dB/km, FSO link length 1 km

Fig. 5 shows the eye diagram pro FSO link length 1 km and attenuation coefficient α_{fog} = 1.61 dB/km. This attenuation coefficient represents relatively pure atmosphere [7]. The results of BER, ISI and jitter are practically the same as in previous Fig. 4. You can see the amplitude of the eye diagram is smaller.

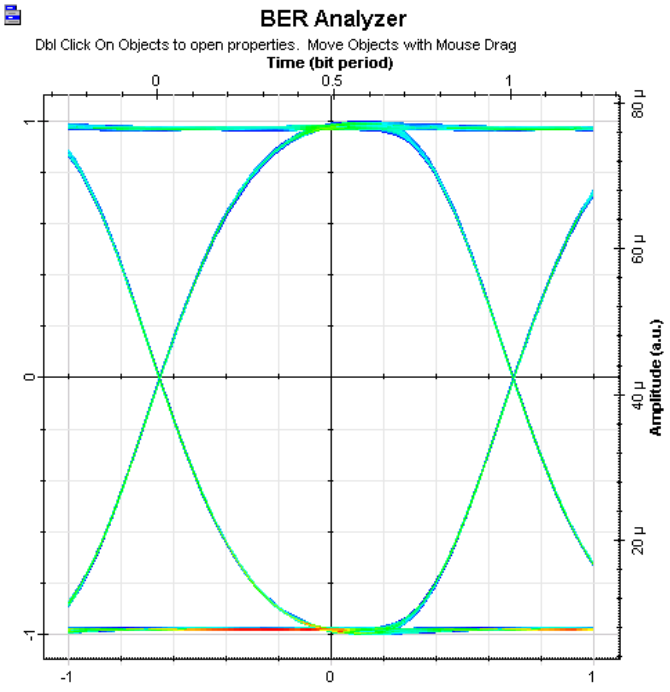


Figure 5. Eye diagram, α_{fog} = 1.61 dB/km, FSO link length 1 km

Fig. 6 shows the eye diagram pro FSO link length 1 km and attenuation coefficient α_{fog} = 13.67 dB/km. This attenuation coefficient represents a haze [7]. The eye opening is narrower, the BER increased a few. ISI and jitter are greater. The amplitude of the eye diagram is 20 times smaller in comparison to Fig. 4.

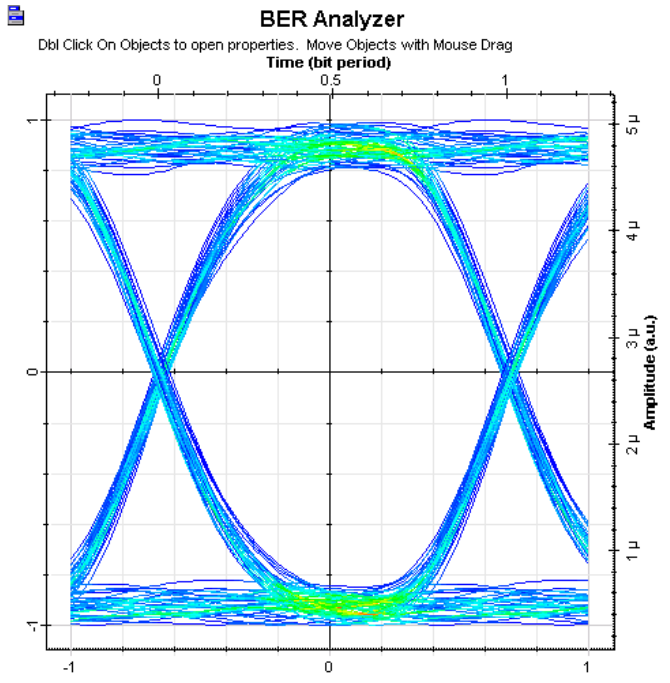


Figure 6. Eye diagram, α_{fog}=13.67 dB/km, FSO link length 1 km

Fig. 7 shows the eye diagram pro FSO link length 1 km and attenuation coefficient α_{fog} = 34.00 dB/km. This attenuation coefficient represents medium fog. The eye diagram broke up, the FSO link can not work.

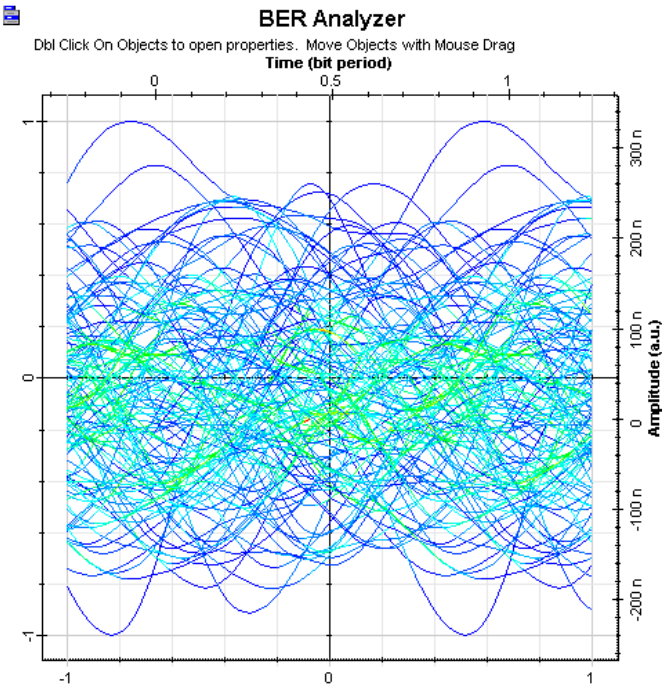


Figure 7. Eye diagram, α_{fog} = 34.00 dB/km, FSO link length 1 km

If we compare from Fig. 4 to Fig. 7, we find out that eye width decreased until the eye diagram broke up. The reason was to great attenuation on the link.

To determine the Q-factor and BER dependence on the attenuation coefficient it was set more attenuation coefficient α_{fog} values, from 0.1 dB/km to 40 dB/km. The measured results are summarized in Tab. IV. FSO link length was 1 km. The results are displayed in Fig. 8.

TABLE IV. Q-FACTOR AND BER VALUES DEPENDENCE ON ATTENUATION COEFFICIENT A_{FOG}

FSO link length 1 km		
α_{fog} [dB/km]	Q-factor	BER
0.1	323.5	0
0.5	303.0	0
1.0	278.7	0
1.5	256.0	0
2.0	234.6	0
5.0	134.6	0
8.0	73.6	0
10.0	48.4	0
12.0	31.8	2.8e-223
15.0	16.9	3.8e-64
16.0	13.5	5.5e-42
17.0	10.8	1.4e-27
18.0	8.6	2.9e-18
19.0	6.9	3.0e-12
20.0	5.5	2.2e-08
21.0	4.4	6.7e-06
22.0	3.5	2.7e-04
23.0	2.7	3.0e-03
25.0	0.0	1
30.0	0.0	1
35.0	0.0	1
40.0	0.0	1

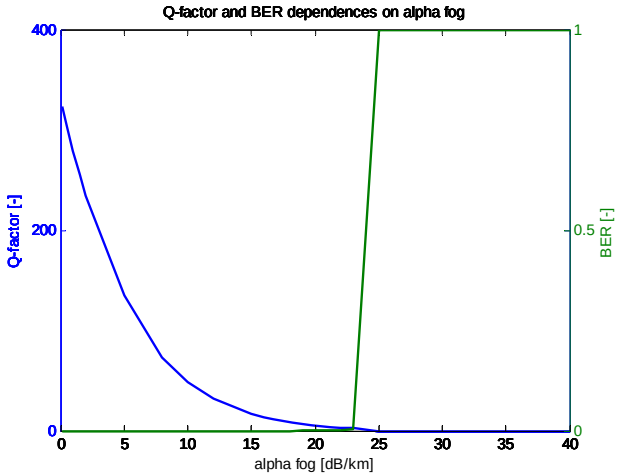


Figure 8. Q-factor and BER dependences on attenuation coefficient α_{fog}

To determine the Q-factor and BER dependence on the FSO link length by attenuation coefficient α_{fog} = 8 dB/km it was set lengths in range from 50 m to 2.4 km. The measured Q-factor and BER results are summarized in Tab. V. for different FSO link lengths. The results are displayed in Fig. 9.

TABLE V. Q-FACTOR AND BER VALUES DEPENDENCE ON FSO LINK LENGTH L

Attenuation coefficient α_{fog} = 8 dB/km		
l [km]	Q-factor	BER
0.05	1494.6	0
0.1	1287.0	0
0.2	944.1	0
0.4	503.5	0
0.6	265.9	0
0.8	139.5	0
1.0	73.6	0
1.2	39.6	0
1.4	22.4	4.4e-111
1.6	12.6	1.2e-36
1.8	7.1	4.6e-13
2.0	4.1	2.1e-05
2.2	2.4	8.7e-03
2.4	0.0	1

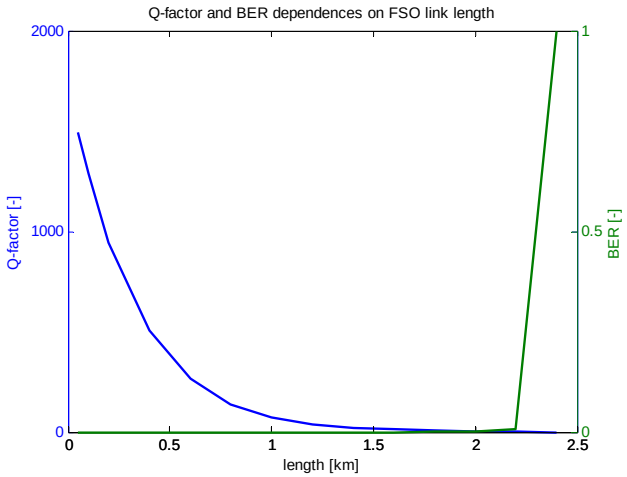


Figure 9. Q-factor and BER dependences on FSO link length l

VI. CONCLUSION

Using software OptiSystem it was simulated atmosphere influence, concretely fog influence, on quality of received signal. The quality is represents by bit error ratio BER. The simulations were run for fixed FSO link parameters, it was changed only the attenuation coefficient by constant link length, in other case the length link was changed by constant attenuation coefficient.

For invariable FSO link length, variable attenuation coefficient, the results displayed in Graph 1 show that BER is relatively long zero and the link works correct. Around value α_{fog}.= 20 dB/km the BER grows up. BER still grows up and the link becomes functionless. Attenuation coefficient α_{fog}.= 20 dB/km it means medium fog [7].

For invariable attenuation coefficient, variable FSO link length, the results displayed in Graph 2 show that for given attenuation coefficient is BER long zero. Around FSO link length 2.2 km the BER grows up until the link becomes functionless. The results deal for above defined parameters.

The fog influences the FSO link the most. The obtained results say that the communication is possible by haze. From the middle fog it appears degradation.

ACKNOWLEDGMENT

This project has been supported by a Czech Science Foundation GA102/09/0550 and SP2011/47.

REFERENCES

[1] *National Weather Service* [online]. 2010 [cit. 2011-04-10]. Terms used by meteorologists, forecasters, weather observers, and in weather forecasts. Available from WWW: <http://www.erh.noaa.gov/er/box/glossary.htm>.

[2] VITÁSEK, Jan, et al. Fog model for Free Space Optics Link. In *Knowledge in Telecommunication Technologies and Optics 2010 KTTO 2010* [online]. 1st edition. Ostrava : VSB-Technical University of Ostrava, 2010 [cit. 2011-04-26]. Available from WWW: <http://ktto.osanet.cz/proceeding.php>. ISBN 978-80-248-2330-0.

[3] SHEIKH MUHAMMAD, S.; KÖHLDORFER, P.; LEITGEB, E. Channel Modeling for Terrestrial Free Space Optical Links. In *7th International Conference on Transparent Optical Networks* [online]. Barcelona : IEEE, 2005 [cit. 2010-11-16]. Available from WWW: < http://ieeexplore.ieee.org/iel5/10062/32275/01505832.pdf?arnumber=150583>.

[4] DORDOVÁ, L. *Method of Atmospheric Transmission Media Characteristics Determination in Optical Spectrum.* Brno: Brno University of Technology, Faculty of Electrical Engineering and Communication, 2009. 137 p. Supervisor prof. Ing. Otakar Wilfert, CSc.

[5] TEJKAL, Vladimír, et al. Dvoustavové modulační formáty v optických přístupových sítích. *Advances in Electrical and Electronic Engineering* [online]. 2010, 8, 4, [cit. 2011-04-26]. Available from WWW: <http://advances.utc.sk/journal2010.html#number4>. ISSN 1804-3119.

[6] *OptiSystem : Component Library*. Kanada : OptiSystem, 2010. 1970 s.

[7] WILFERT, Otakar. *Fotonika a optické komunikace - přednášky*. Brno : Vutium, 2000. 122 s. ISBN 978-80-214-3537-7.

Telecommunication Systems

chairmen:

Gokhan ILK and Bohdan BOROWIK

FPGA Implementation of Turbo Decoders Using BCJR Algorithm

Onur ATAR*, Murat H. SAZLI, H. Gokhan ILK

*TUBITAK Uzay, Space Technologies Research Institute, Ankara TURKEY
onur.atar@uzay.tubitak.gov.tr
Ankara University, Faculty of Engineering, Electronics Engineering Department, 06100, Tandogan Ankara, TURKEY
sazli@eng.ankara.edu.tr, ilk@ieee.org

***Abstract*—The most challenging design issue for turbo codes, which is a successful channel coding method to approach the channel capacity limit, is the design of the iterative decoders which perform calculations for all possible states of the encoders. BCJR (MAP) algorithm, which is used for turbo decoders, embodies complex mathematical operations such as division, exponential and logarithm calculations. Therefore, BCJR algorithm was avoided and the sub-optimal derivatives of this algorithm such as Log-MAP and Max-Log-MAP were preferred for turbo decoder implementations. BCJR algorithm was reformulated and wrapped into a suitable structure for FPGA implementations in previous works. Reformulated BCJR algorithm is implemented in this work. Complex mathematical operations which run slowly on hardware (division, exponential and logarithm calculations) are read from look-up-tables and high performance calculation structures are established. Implemented system is verified through simulations. It is observed that the BER performance of the proposed algorithm is better than the Log-MAP algorithm as expected.**

Keywords: MAP algorithm, Turbo codes & coding, real time embedded programming

I. INTRODUCTION

Shannon's 1948 paper entitled "A Mathematical Theory of Communication" [2] has been considered as the birth of a new field, "error control coding". In that paper Shannon defined the concept of "channel capacity". He then showed that there exist error control codes that can yield arbitrarily low errors at the receiver output, so long as the transmission rate through the channel is less than the channel capacity. Although he showed the existence of such codes, he did not specify how to construct them.

Even though many efficient coding and decoding schemes have been developed following Shannon, not until the invention of the state-of-the-art turbo codes in 1993 have we been able to approach to the channel capacity limit within just a few tenths of a dB. In other words, turbo codes have closed the significant gap between the coding gains so far achieved using the conventional coding and decoding schemes, and the channel capacity limit.

In order to implement an efficient turbo decoder, a suitable decoding algorithm has to be chosen. Turbo codes have been originally implemented with BCJR (Bahl, Cocke, Jelinek, Raviv) [4] algorithm. However, this algorithm performs complex mathematical operations such as multiplication, division and logarithmic calculations. Therefore, engineers have avoided implementing this complex algorithm and preferred the sub-optimal derivatives of the BCJR (MAP) algorithm such as the Log-MAP and the Max-Log-MAP algorithms which are much simpler to implement but yield worse BER performances [5].

With the technology we had in the early 90s when the turbo codes were invented, it was impossible to implement a complex structure such as the BCJR algorithm on a single IC. However, today's VLSI technology allows us to implement not only a channel coding algorithm but also a whole communication system on a single IC. Therefore, turbo decoders using the BCJR algorithm which have been avoided in spite of its superior BER performance became implementable on modern FPGAs.

II. TURBO CODING

"Error control coding" or "channel coding" is an essential part of a communication system. By adding redundancy to the message to be transmitted in a systematic manner, it is possible to detect and correct errors which unavoidably occur during the transmission through the

channel communication. “Turbo coding” is a channel coding structure and it is the base of this work. This section presents theoretical background of turbo coding.

A. Turbo Encoder

A turbo encoder consists of a parallel concatenation of two (or more) systematic codes. Each of the encoders/decoders is called as “component encoders/decoders”. If there are two component encoders in the turbo encoder, it is referred to as a “two dimensional turbo code”.

Originally, turbo codes were developed with RSC (Recursive Systematic Convolutional) encoders. It is pointed out in [3] that the bit error rate of classical NSC (Non-Systematic Convolutional) code is lower than that of a classical Systematic code with the same code memory at large signal-to-noise ratio (SNR). However, at low SNR it is in general the other way around. They also introduced in [3] that RSC codes can be better than the best NSC code at any SNR for high code rates.

B. Recursive Systematic Convolutional Encoders

Before we present the RSC, let us first review classical Non Systematic convolutional encoder. Consider a binary rate R=1/2 convolutional encoder with constraint length K and memory v=K-1. The input to the encoder at time k is a bit d_k and the corresponding codeword C_k is the binary couple (X_k, Y_k) which are calculated as given in eq. 1 and 2 respectively.

$$X_k = \sum_{i=0}^{K-1} g_{1i} d_{k-i} \quad \text{mod.2} \quad g_{1i} = 0,1 \tag{1}$$

$$Y_k = \sum_{i=0}^{K-1} g_{2i} d_{k-i} \quad \text{mod.2} \quad g_{2i} = 0,1 \tag{2}$$

where $G_1 : \{g_{1i}\}$, $G_2 : \{g_{2i}\}$ are two code generators, generally expressed in octal form. NSC defined by code generators $G_1 = 17$, $G_2 = 11$, with memory v=3 is given in Figure 1.

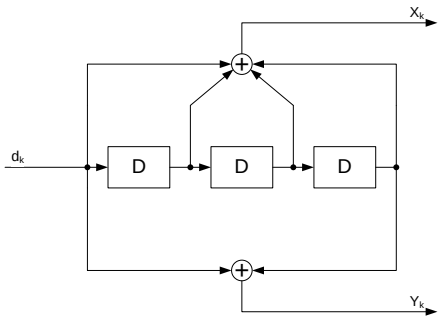


Fig. 1. Classical NSC encoder

RSC code obtained from the NSC using a feedback loop is given in Figure 2. Notice that one of the two outputs (in this case $X_{k)}$ is set equal to the information bit d_k, therefore making the code systematic. Also notice that for an RSC code, the shift register (memory) input is no longer the information bit d_k but is a new variable a_k which is recursively calculated according to eq. 3.

$$a_k = d_k + \sum_{i=1}^{K-1} \gamma_i a_{k-i} \quad \text{mod.2} \tag{3}$$

where γi is respectively equal to g_{1i}, if $X_k = d_k$ and to $g_{2i,}$ if $Y_k = d_k$.

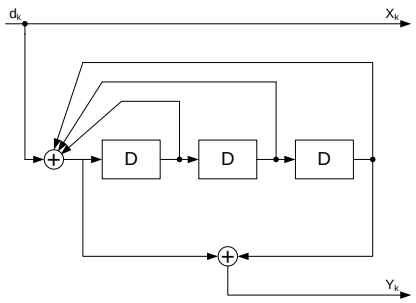


Fig. 2. RSC encoder (rate 1/2)

C. Parallel Concatenation of RSC Encoders

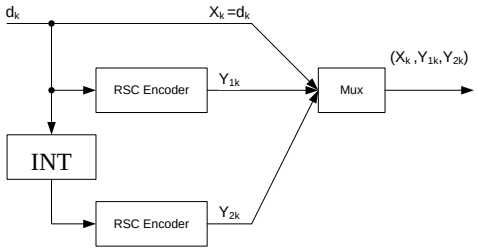


Fig. 3. Turbo encoder

In Figure 3, a rate 1/3 turbo encoder is given. Data (d_1, d_N) is encoded with blocks. Block size is determined by the size of the interleaver. Original data stream is encoded by the first encoder and produces a parity bit sequence (Y_{11}, Y_{1N}). The interleaver shuffles original data stream and then this scrambled version of the data stream is encoded by the second encoder producing the second parity bit sequence (Y_{21},Y_{2N}). Interleaving may be according to a certain rule, or it may be totally random. Usually the bigger is the interleaver size; the better is the performance of the turbo decoder.

Parity bits of the both component encoders may be multiplexed with the information bit and transmitted. This yields a rate 1/3 turbo code. Alternatively, parity bits of the component encoders may be multiplexed with the information bit according to a pre-defined pattern. This is called as "code puncturing" and the resultant code is called a "punctured code". For instance, (X_k, Y_{1k}) is transmitted at one signaling interval, following that (X_{k+1}, $Y_{2(k+1)}$) is transmitted. Therefore, for each information bit, only one parity bit is transmitted, and that results a rate 1/2 turbo code. Since the parity bit of one component decoder is deleted at each signaling interval, corresponding decoder uses "zero" for that interval during decoding.

D. Turbo Decoder

To facilitate the following detailed explanation of the underlying concepts of turbo decoder and the iterative decoding performed by it, we first present the turbo decoder in its original form, which has been introduced in [3], in Figure 4.

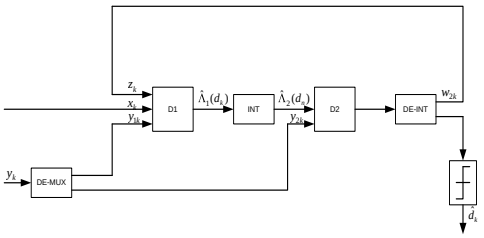


Fig. 4. Turbo decoder

For the first iteration of the turbo decoder, z_k's are initialized to zero. Received bits x_k's and y_k's are multiplexed to get the appropriate parity bits for the component decoders. Parity bits are provided to the corresponding decoders according to the puncturing pattern used in the turbo encoder. If the turbo code is unpunctured, then at each time index k y_{1k} and y_{2k} are provided to decoders D1 and D2 respectively.

D1 uses x_k and y_{1k} and produces a "soft" estimate of the decoded bit d_k. A relevant piece of information is extracted from this soft estimate (which is called as "extrinsic information") and passed on to D2 after proper interleaving (same interleaver which is used in encoder). D2 uses this extrinsic information along with x_k and y_{2k}, in turn yielding a better estimate of bit d_k. Then the extrinsic information produced by D2 is passed on to D1 for the next iteration to improve the estimates on all d_k's in the block. Until a satisfactory BER is achieved, this iterative process continues. Usually, less than 20 iterations yield a BER of 10^{-5}.

After a predetermined number of iterations is reached, soft output of D2 is passed on to a hard limiter (with a threshold set to zero) to obtain the estimates on the information bits.

III. REFORMULATION OF THE BCJR ALGORITHM

In this section we will reformulate the BCJR algorithm via some matrix manipulations [6]. In the following we will consider a recursive convolutional encoder with a constraint length K and code memory v = K – 1. There are 2^v states of this encoder. We also suppose that BPSK modulation is used, i.e. bit one is mapped to +1, and bit zero is mapped to -1.

A. Calculation of the Forward Metrics (Alpha Coefficients)

Let us begin with the recursive equation to obtain the α coefficients of the BCJR algorithm according to eq. 4.

$$\alpha_k(m) = \frac{\sum_{m'}\sum_{i=-1}^{i=+1}\alpha_{k-1}(m')\gamma_i(R_k,m',m)}{\sum_{m}\sum_{m'}\sum_{i=-1}^{i=+1}\alpha_{k-1}(m')\gamma_i(R_k,m',m)} \quad (4)$$

where m = 0, 1, 2, …, M, is the index of the states with $M = 2^v - 1$.

After some derivations defined in [6] and [7], the reformulation of this recursive equation is turned into an implementable structure of the alpha calculator which is given in Figure 5.

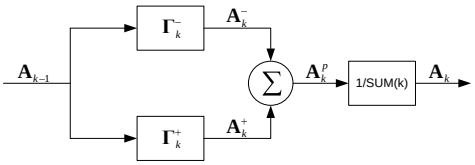


Fig. 5. The structure of the alpha calculator

B. Calculation of the Backward Metrics (Beta Coefficients)

As given in eq. 5, a similar procedure was followed to formulate the β coefficients of the BCJR algorithm in matrix notation in [6] and [7].

$$\beta_k(m) = \frac{\sum_{m'}\sum_{i=-1}^{i=+1}\beta_{k+1}(m')\gamma_1(R_{k+1},m,m')}{\sum_{m}\sum_{m'}\sum_{i=-1}^{i=+1}\alpha_k(m)\gamma_1(R_{k+1},m,m')} \quad (5)$$

After some derivations defined in [6] and [7], the reformulation of this recursive equation is turned into an implementable structure of the beta calculator which is given in Figure 6.

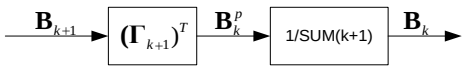


Fig. 6. The structure of the beta calculator

C. Calculation of the Logarithmic Likelihood Ratios (LLRs)

Now, we will show how we can compute the logarithm of the likelihood ratios (LLR) associated with each bit d_k using the previously computed $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{\Gamma}$ matrices. LLR associated with each bit d_k is calculated as given in eq. 6.

$$\Lambda(d_k) = \ln \frac{\sum_m \sum_{m'} \gamma_{+1}(R_k, m', m)\alpha_{k-1}(m')\beta_k(m)}{\sum_m \sum_{m'} \gamma_{-1}(R_k, m', m)\alpha_{k-1}(m')\beta_k(m)} \quad (6)$$

After some derivations defined in [6] and [7], the reformulation of this recursive equation is turned into an implementable structure of the LLR calculator which is given in Figure 7.

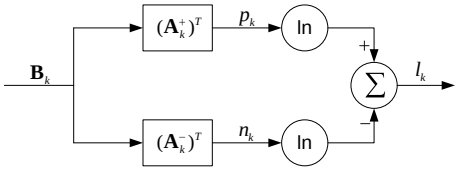


Fig. 7. The structure of the LLR calculator

IV. FPGA IMPLEMENTATION OF TURBO DECODERS USING BCJR ALGORITHM

Turbo decoders are highly configurable systems. Some of the configuration parameters are chosen by the design engineer, some of them are defined by the 3GPP standard [8]. The parameters used in this work and their values are given in Table 1.

Table 1. Design parameters

Parameter	Value
Block Length	128-4096
# of RSC Encoders	2
Constraint Length of RSC Encoders	4
Generator matrix of RSC Encoders	G={7,5}(oktal)
Code Rate	R=1/3
Code Puncturing	None
Decoding Algorithm	BCJR (MAP)
Iteration #	1-20

A. General Hardware Structure of the System

General hardware structure of the implemented system is given in Figure 8. The received data stream $R_k(x_k, y_{1k}, y_{2k})$ is handled by the "input handling system" in order to be stored and provided to the appropriate decoders.

As indicated before, reformulated BCJR algorithm is used in component decoders [6]. The hardware structure of this algorithm is given in Figure 9.

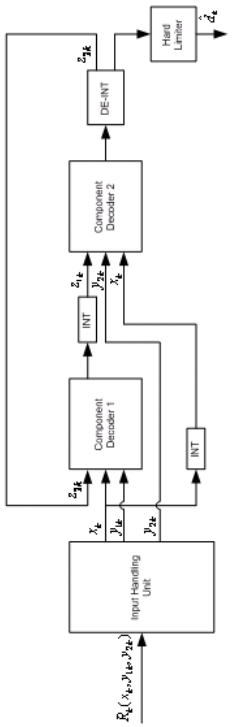


Fig. 8. General hardware structure of the system

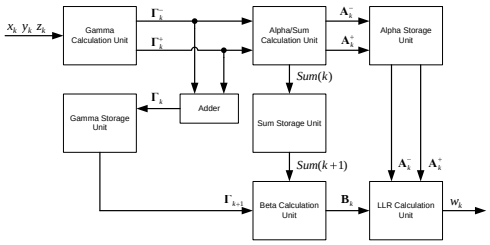


Fig. 9. Hardware structure of the component decoders

Coefficients used by the BCJR algorithm and the noisy received data stream are defined as real numbers. Therefore, these numbers must be represented as "fixed-point" on hardware. Nine-bit numbers are used for both the received data stream and the internal coefficients. The most significant bit is the sign bit- bit "1" for negative and bit "0" for positive. Next significant 3 bits are the integer part of the

number. For the fractional part of the number, the least significant 5 bits are used.

B. Gamma Calculation Unit

The state transition metrics (gamma coefficients) is calculated as given in eq. 7.

$$\gamma_i(R_k, m', m) = \frac{1}{2\pi\sigma^2}\frac{\exp(z_k/2)}{1+\exp(z_k)}\exp(iz_k/2) \cdot q(d_k = i \mid S_k = m, S_{k-1} = m') \cdot \exp\{\frac{-1}{2\sigma^2}[(x_k - i)^2 + (y_k - Y_k)^2]\} \quad (7)$$

Eq. 7 can also be considered as product of three terms. The $\frac{1}{2\pi\sigma^2}\frac{\exp(z_k/2)}{1+\exp(z_k)}\exp(iz_k/2)$ term is the first term, the $q(d_k = i \mid S_k = m, S_{k-1} = m')$ term is the second term and the $\exp\{\frac{-1}{2\sigma^2}[(x_k - i)^2 + (y_k - Y_k)^2]\}$ term is the third term of this product. These three terms are multiplied via the digital elements on hardware. Constants and exponential expressions in these terms are not calculated but read from look-up tables which results in reduced usage of hardware elements (except block memories).

C. Alpha, Beta and LLR Calculation Units

These units implement not only multiplication and addition operations but also division operation which runs slowly on hardware despite the high speed modern FPGAs. Therefore, the division operation is implemented via look-up tables.

Logarithmic calculation is the only complex calculation implemented in the LLR unit. Therefore, it is also implemented via look-up tables similar to alpha and beta calculation units.

D. BER Performance of the BCJR Turbo Decoder

BCJR turbo decoder implemented at this work yields better BER performance than the Xilinx Log-MAP turbo decoder [9]. BER performance comparison between these two turbo decoders is given in Table 2.

E. Total Throughput Performance of the BCJR Turbo Decoder

As indicated before, turbo decoding operation is an iterative operation. For every data block, more than one iteration is completed by the decoder. Thus, turbo decoders yield better BER performance. However, these iterations decrease the total throughput of the decoder. Total throughput performance comparison between the BCJR and the Log-MAP turbo decoders is given in Table 3.

Table 2. BER performance comparison of R=1/3 BCJR turbo decoder and the Log-MAP turbo decoder after 5 iterations

Block Length	512 Bits		1024 Bits		2048 Bits	
SNR	BCJR	Log-MAP	BCJR	Log-MAP	BCJR	Log-MAP
SNR = 0.5 dB	3.10^{-3}	4.10^{-2}	4.10^{-4}	3.10^{-2}	10^{-4}	2.10^{-2}
SNR = 1 dB	8.10^{-4}	10^{-2}	2.10^{-5}	2.10^{-3}	6.10^{-6}	6.10^{-4}
SNR = 1.5 dB	2.10^{-5}	10^{-3}	10^{-6}	4.10^{-5}	4.10^{-7}	2.10^{-6}
SNR = 2 dB	2.10^{-6}	2.10^{-4}	$<10^{-7}$	10^{-6}	$<10^{-7}$	10^{-7}

Table 3. Total throughput performance comparison of R=1/3 BCJR turbo decoder and the Log-MAP turbo decoder at 139 MHz and 349 MHz clock frequency respectively (Mbps)

Block Length	512 Bits		1024 Bits		2048 Bits	
Iteration #	BCJR	Log-MAP	BCJR	Log-MAP	BCJR	Log-MAP
Iteration = 3	7,87	32,93	7,88	35,58	7,90	37,96
Iteration = 5	4,69	20,68	4,72	22,42	4,74	24,00
Iteration = 7	3,35	15,08	3,36	16,37	3,38	17,54
Iteration = 9	2,61	11,86	2,62	12,89	2,63	13,82

Table 4. Decoding latency comparison of R=1/3 BCJR turbo decoder and the Log-MAP turbo decoder (clock cycles)

Block Length	512 Bits		1024 Bits		2048 Bits	
Iteration #	BCJR	Log-MAP	BCJR	Log-MAP	BCJR	Log-MAP
Iteration = 3	10878	4664	21630	8632	43134	16184
Iteration = 5	17102	7424	33998	13696	67790	25600
Iteration = 7	23326	10184	46366	18760	92446	35016
Iteration = 9	29550	12944	58734	23824	117102	44432

F. Decoding Latency of the BCJR Turbo Decoder

Turbo codes are block codes. Therefore, turbo decoders perform decoding operations on input blocks and output the decoded data

blocks. The time interval between the reception of the first bit of a block and the output of the first decoded bit of the same block is called "decoding latency".

Decoding latency comparison between the BCJR and the Log-MAP turbo decoders is given in Table 4.

G. Implementation Report of the BCJR Turbo Decoder

BCJR algorithm is a computationally complex algorithm. Implementing complex mathematical operations such as multiplication and division significantly increases the usage of hardware elements. Therefore, BCJR algorithm uses more hardware elements and runs slower on hardware due to its complex mathematical operations. The impact of this disadvantage is reduced by using look-up table for complex operations except multiplication. Implementation reports of the BCJR turbo decoder and the Xilinx Log-MAP turbo decoder which operate on exactly the same platform are given in Table 5 and 6 respectively.

Table 5. Implementation report of the R=1/3 rate BCJR turbo decoder

Preferences		
Xilinx FPGA	XC6VLX75T	
LUT/FF Pairs	2264	
Slice LUT	36254	
Slice Register	2404	
Blok RAM (36k)	142	
Blok RAM (18k)	0	
DSP Blocks	0	
Speed Grade	-1	-3
Maximum Clock Frequency	130 MHz	139 MHz

Table 6. Implementation report of the R=1/3 rate Log-MAP turbo decoder

Preferences		
Xilinx FPGA	XC6VLX75T	
LUT/FF Pairs	3765	
Slice LUT	3712	
Slice Register	4062	
Blok RAM (36k)	6	
Blok RAM (18k)	7	
DSP Blocks	0	
Speed Grade	-1	-3
Maximum Clock Frequency	285 MHz	349 MHz

V. CONCLUSION

The turbo decoding structure based on a previous work [6] is implemented in this work. As indicated in the introduction section, the BCJR turbo decoder is compared with the Xilinx Log-MAP turbo decoder [9]. It is observed that the BCJR turbo decoder yields a better BER performance than the Xilinx Log-MAP turbo decoder as expected.

In spite of its superior BER performance, implementation of the BCJR algorithm has been avoided because of its complexity considering the past VLSI technology. However, modern VLSI technology allows us to implement this algorithm at reasonable costs. The prospective application areas of our proposed implementation are:

- Applications that require low BER with a disclaimed throughput performance.
- Power constraint applications where the desired BER is claimed at low SNR which means low power consumption.
- Applications where both throughput and BER are important design issues. In such a case the proposed approach can be used in parallel by using multiple turbo decoding engines which can provide very high throughput at an already provided low BER. This cannot be achieved by Xilinx Log-MAP turbo decoder approach.

REFERENCES

1. Sazlı, M., H., Neural Network Applications to Turbo Decoding, Ph. D. Dissertation, Syracuse University, 2003.
2. Shannon, C. E., "A Mathematical Theory of Communications", Bell System Technical Journal, Vol. 27, pp.379-423, 623-656, 1948.
3. Berrou, C., Glavieux, A., Thitimajshima, P., "Near Shannon Limit Error-Correcting Coding and Decoding: Turbo Codes", Proceedings of IEEE International Conference on Communication, pp. 1064-1070, 1993.
4. Bahl, L. R., Cocke, J., Jelinek, F., Raviv, J., "Optimal Decoding of Linear Codes for Minimizing The Symbol Error Rate", IEEE Transactions on Information Theory, Vol. 20, pp. 284-287, 1974.
5. Robertson, P., Hoeher P., "Optimal and Sub-Optimal Maximum a Posteriori Algorithms Suitable for Turbo Decoding", European Transactions on Telecommunications, Vol. 8, pp. 119-125, 1997.
6. Sazlı, M., H., "Neural Network Implementation of BCJR Algorithm Based on Reformulation Using Matrix Algebra", IEEE International Symposium on Signal Processing and Information Technology, pp. 832-837, 2005.
7. Sazlı, M., H., "Neural Network Implementation of BCJR Algorithm", Digital Signal Processing, Elsevier, Vol 17; pp. 353-359, 2007.
8. 3GPP, "3GPP Technical Specification", http://www.3gpp.org, 2010.
9. Xilinx, Inc., "3GPP Turbo Decoder v4.0 Product Specification", Technical Journal, http://www.xilinx.com, 2009.

Significantly Asymmetric Spectral Profile for a Very High-bit-rate Digital Subscriber Line

Jiri Vodrazka, Petr Jares
Czech Technical University in Prague, Faculty of Electrical Engineering
Department of Telecommunication Engineering
Prague, Czech Republic
vodrazka@fel.cvut.cz, jaresp@fel.cvut.cz

***Abstract*—The very high speed digital subscriber line (VDSL) is standardized for the frequency bands 12, 18 and 30 MHz. The spectral profiles are characterized by alternating bands for upstream and downstream directions when using frequency duplex. This paper describes a specific variant of choice for access network in the Czech Republic under the terms of spectrum released by Telefónica O2 CR. Further discussion achievable bit rates under different conditions and by the expected bit rates profiles, which will be offered to subscribers. The resulting bit rates are calculated depending on the length of the line with a copper core diameter of 0.4 mm and the number of disturbers for upstream and downstream direction. Higher data rates can be achieved with vectored modulation VDMT for FEXT crosstalk cancellation.**

Keywords-digital subscriber line; access network; VDSL2

I. INTRODUCTION

While in some parts of the world for a long time to install all-optical lines FTTH (Fiber to the Home) based on a passive optical network (PON), the FTTH in the Czech Republic is progressing very slowly. In May 2011, can expect at least increase the existing rate of metallic lines, as shown in terms of spectrum management, supplemented by Telefónica O2 CR [2].

The very high speed digital subscriber line (VDSL) has been standardized for the frequency bands first 12, then almost 18 and then 30 MHz [1]. The spectral profiles are characterized by alternating bands for upstream and downstream directions over a frequency division duplex (FDD). The crossover frequencies are different for a plan called 997, suitable for transmission speed and symmetry of 998, suitable for asymmetric data rates [3]. For Europe, according to the ITU-T G.993.2 both plans may be the choice of operator. There is, however, can be combined in one network because of the near end crosstalk (NEXT).

The symmetry of bit rates is suitable for upload the subscriber files to servers, remote access via VPN, etc. The asymmetry of bit rates suited for services such as IPTV and video on demand. It is also suitable for network operators from marketing point of view because the asymmetry line has one of a speed significantly higher than either speed symmetrical line (the sum of the rate is same in both cases).

II. SECOND GENERATION OF VDSL LINES

The second generation of VDSL2 systems (according to ITU-T G.993.2 Recommendation) admits only one modulation, which is DMT (MCM). Other substantial changes against the original VDSL specifications come to VDSL2 from ADSL2, especially Trellis Code Modulation (TCM) and the ability to correct individual errors in the receiver unit. The error correction using RS code, as well as interleaving is obvious for all considered types of connections. Similarly to ADS2, it is possible to adapt the transmission speed during the operation (SRA – Seamless Rate Adaptation) and control the transmit power in order to reduce the crosstalk to the neighboring pairs (sleep mode).

In addition to the existing frequency bands, VDSL2 introduces two more ones (DS3 and US3, in some frequency plans also DS4), thus offering exploitation of frequencies up to 30 MHz. The individual profiles are specified in the Annexes of ITU-T G.993.2 Recommendation, accepting the basic frequency plans that have been called since the first generation of VDSL 997 (more suitable for symmetric applications) and 998: Annex A – for North America – plan 998, Annex B – for Europe – plans 997 and 998, and Annex C – for Japan – plan 998.

For Europe and North America there have been defined frequency plans up to 12 MHz, while for Japan up to 30 MHz. The basic alternatives are illustrated in Figure 1. The number of used frequency bands in the downstream (DS1 to DS3) and an upstream (US0 to US3) direction is determined by the required transmission speeds in both directions and also by the length and attenuation of the telecommunication line. Even finer classification of the arrangement of subscriber lines is based on the use of the lowest band for upstream direction (US0); this band is either unused or it is used similarly as in ADSL2 in connection with possible operation of analog telephone (POTS) or ISDN line. There is a set of PSD masks defined for use in Europe, which are differentiated according to 997 or 998 frequency plans.

The plan 998 is the original plan with maximal frequency 12 MHz. The plan 998E17 is plan 998 directly extended to maximal frequency 17.664 MHz. The plan 998E30 is plan 998 directly extended to maximal frequency 30 MHz.

III. FREQUENCY PLAN FOR THE CR LINES IN THE O2 NETWORK

The selection of the VDSL spectral profile is very important strategic result for access network provider. The spectral profile 998ADE17 was chosen for VDSL2 lines operating in the access network owned by Telefonica O2 CR marked, as is clear from the document Spectrum management [2]. It is characterized by an increased asymmetry of bands since the 12 MHz frequency band is intended solely for the downstream direction, as shown on Fig. 1 and Tab. 1. This corresponds to the frequency plan (spectral mask) B8-12, which are characterized by the lowest band US0, and compatibility with basic ISDN (VDSL2 over ISDN).

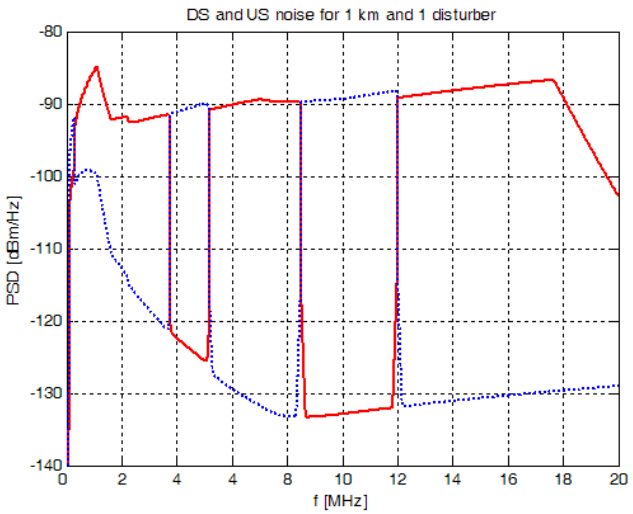


Figure 2. Power spectral density of noise for profile 998ADE17, length 1 km and 1 disturber in the cable (downstream – red solid line, upstream – blue doted line)

Used frequency bands are evident from Fig. 2, which shows the noise power spectral density (PSD) with dominant far end crosstalk FEXT, which is reflected in both directions of transmission in each zone, which is used for the useful signal. Specifically, the noise is calculated for the connection length of 1 km and 1 disturber.

Based on the observed and calculated noise power spectral density at the receiver input signal can be estimated achievable bit rate for different cases.

IV. ESTIMATED TRANSMISSION RATE

Estimation of achievable bit rate in frequency-dependent channel with crosstalk and DMT modulation can be found on the calculations in a number of sources, such as [4], [8], [10] which can be derived from the simplified relations (1) and (2). First, the signal to noise ratio for the i-th sub-channel can be calculated:

$$SNR_i = \frac{10^{-\frac{\alpha_i \cdot d}{10}}}{\frac{P_{AWGN}}{P_i} + n^{0.6}\left(K_{NEXT}\cdot\left(i\frac{B}{N}\right)^{3/2} + K_{FEXT}\cdot d\cdot\left(i\frac{B}{N}\right)^{2}\cdot 10^{-\frac{\alpha_i \cdot d}{10}}\right)} \quad (1)$$

where

α_i is propagation constant in dB/km
d is length of line in km
P_{AWGN} is the mean power of noise in sub-channel
P_i is the mean power of signal in sub-channel i
n is number of VDSL2 lines in the same cable
K_{NEXT} is constant for near end crosstalk (worst case in the cable)
K_{FEXT} is constant for far end crosstalk (worst case in the cable)
B is total frequency band
N is number of sub-channels

The real used sub-channels must be selected for band US0, 1, 2 for calculation of upstream and DS1, 2, 3 for downstream.

TABLE I. Cutoff frequency for profile 998ADE17

Band	US0		DS1		US1		DS2		US2		DS3	
f [kHz]	120	276	276	3750	3750	5200	5200	8500	8500	12000	12000	17664

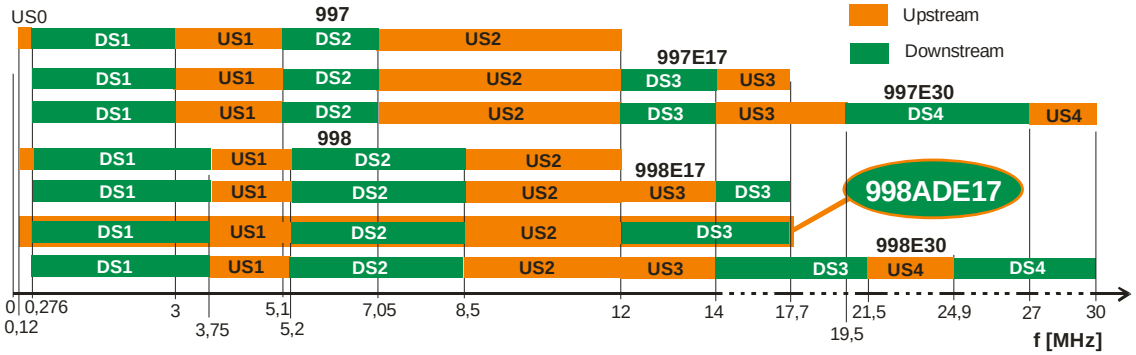


Figure 1. Some spectral profiles of VDSL2 lines for Europe – for CR selected profile 998ADE17 is marked

The total bit rate can be calculated by summing up the bit rates in the sub-channels:

$$BR = \frac{2 \cdot B}{N_{CP} + 2 \cdot N} \cdot \sum_{i=1}^{N} \log_2 \left(1 + \frac{SNR_i}{K_b} \right) \quad \text{[bps]} \qquad (2)$$

where
N_{CP} is length of cyclic prefix for DMT modulation
K_b is Shannon gap for required bit error rate

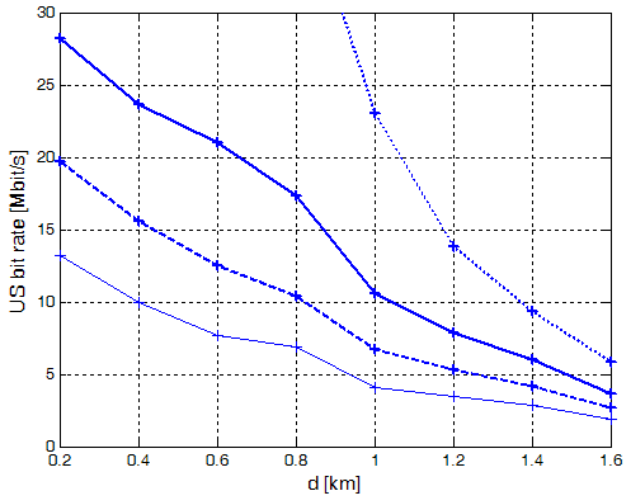


Figure 3. Upstream bit rates length dependence for profile 998ADE17 and number of disturbers: single active line on the local cable (dotted line), worst case for one disturber (full bold line), 20% penetration of the same type of lines (dashed line) and for 100% penetration (thin solid line)

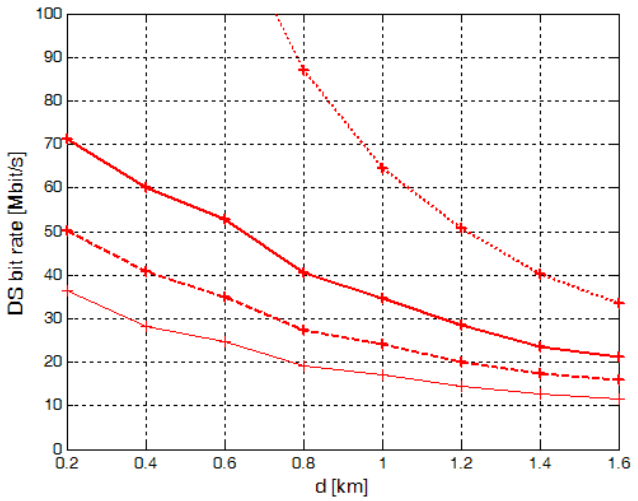


Figure 4. Downstream bit rates length dependence for profile 998ADE17 and number of disturbers: single active line on the local cable (dotted line), worst case for one disturber (full bold line), 20% penetration of the same type of lines (dashed line) and for 100% penetration (thin solid line)

The resulting bit rates are calculated depending on the length of the line with a copper core diameter of 0.4 mm and the number of disturbers for the upstream direction in Fig. 3 and the downstream direction in Fig. 4. It is always plotted the theoretical bit rate for a single connection on the local cable, which does not crosstalk interfere, but only white noise in the background (dotted line), the dependence for the worst case of one disturber (full bold line), the dependence for a 20% penetration of the same type of cable lines (dashed line) and the dependence for 100% penetration of the same type of lines (thin solid line). The calculation of different types of lines and the situation allows on-line simulation program [9]. The frequency profile 998ADE17 corresponds to the frequency plan marked B8-12.

V. Conclusion

It is obvious that with increasing number of active lines in the cable due to crosstalk bit rate decreases rapidly. If we consider that the penetration of active lines will be around 20% can be offered downstream bit rate 36 Mbps and upstream bit rate 12 Mbps for the line to a length of 0.6 kilometers. For the line to a length of 1.6 kilometers will be downstream bit rate 16 and upstream bit rate 5 Mbps. This is the asymmetric transmission rates, but not as much as it is for ADSL2+ (typically downstream 16 Mbps / upstream 0.5 Mbps). However, according to the published wholesale supply company Telefonica O2, it seems that the highest rate offered will be 25.6 / 2 Mbps. Apparently the access service provider keeps a margin for increased speed in the second stage also operate on the basis of connections in real network.

Higher data rates can be achieved with vectored modulation VDMT for FEXT crosstalk cancellation [5], [6], [11] practically feasible as a partial crosstalk cancellation [7]. The bit rate equivalent of single active line on the local cable can be expected for full crosstalk cancellation (dotted line on Fig. 3 and 4).

The VDSL2 lines provide a significant increase in downstream bit rate than ADSL2+ for short lines, the upstream bit rate increase is more pronounced in relative comparison.

Acknowledgment

This paper has originated thanks to the support from The Czech Ministry of Education, Youth and Sports within the project MSM6840770014.

References

[1] Recommendation: ITU-T G.993.2. Very high speed digital subscriber line transceivers 2 (VDSL2).

[2] Reference offer: Zpřístupnění účastnického kovového vedení společnosti Telefónica O2 Czech Republic. Příloha 16: Správa spektra (Aktualizováno 1.11.2010). On-line: http://www.o2.cz/public_part/1b/d/230441_417484_RUO_6._5_Priloha_16___Sprava_spektra_s_VDS_finalL.pdf

[3] Vodrazka, J.: Varianty přípojek VDSL2. Access Server. ISSN 1214-9675. On-line: http://access.feld.cvut.cz/view.php?cisloclanku=2006052401

[4] Brady, M. H., Cioffi, J. M.: The Worst-Case Interference in DSL Systems Employing Dynamic Spectrum Management. In Journal on Applied Signal Processing. Hindawi Publishing Corporation EURASIP. Volume 2006, Article ID 78524, pp. 1-11.

[5] Vodrazka, J.: Multi-carrier Modulation and MIMO Principle Application on Subscriber Lines. In Radioengineering, Vol. 16, No. 1, December 2007. ISSN 1210-2512.

[6] Lafata, P., Vodrazka, J.: Practical Application of FEXT Models for VDMT Modulation. In Proceedings of VIIIth Conference KTTO 2008. Ostrava: VŠB - Technical University of Ostrava, 2008, p. 85-88. ISBN 978-80-248-1719-4.

[7] Cendrillon, R., Ginis, G., Moonen, M., Acker, K.: Partial Crosstalk Precompensation in Downstream VDSL. In Signal Processing 84. Elsevier (2004), pp. 2005-2019.

[8] Starr, T., Cioffi J. M., Silverman, P. J.: Understanding Digital Subscriber Line Technology. The Book. Prentice Hall PTR, Upper Saddle River, USA, Jan. 1999. ISBN 0-13-780545-4.

[9] xDSL simulator. Matlab Server. On-line: http://matlab.feld.cvut.cz/view.php?cisloclanku=2005071801

[10] Róka, R.: Modeling of Environmental Influences at the Signal Transmission by means of the VDSL and PLC Technologies, In: International Journal of Electrical Communication Networks and Information Security – IJCNIS, Vol. 1, No. 1, 2009, pp. 6-13, ISSN 2073-607X

[11] Róka, R.: The Utilization of the Vectored DMT for the FEXT-free Signal Transmission by means of the VDSL Technology, In: IWSSIP 2008 – 15th International Workshop on Systems, Signals and Image Processing, Bratislava (Slovakia), 25. - 28. 6. 2008, pp. 299-302, ISBN 978-80-227-2856-0

Simulation of data transfer from the aerobatic plane

Petr Chlumsky, Zbynek Kocur, Vladimir Machula
Department of Telecommunication
Engineering
Faculty of Electrical Engineering
Czech Technical University in Prague
Prague, Czech Republic
Email: chlumpet@fel.cvut.cz

***Abstract*—This article deals with the optimization of data transmission from the fast-moving objects, typically video signals from aerobatic aircraft. Optimization is performed on the second layer of the RM OSI and uses multiple transmission channels. The simulation was designed in network simulation tool OMNeT++ and is based on the wireless standard IEEE 802.11 (Wi-Fi). For simulation were created two cases using multiple transmission channels. These cases were compared with case of common wireless system which uses a single transmission channel. Proposed cases reached better error rates and proved their usefulness. They are also independent on the transmission technology (due to implementation on the second layer of RM OSI), therefore are versatile.**

***Keywords* - wireless communication; simulation;**

I. Introduction

The High Speed Mobile Communication project is focused on the design of the transmission system that would enable wireless transmission of the video and telemetric data from aerobatic plane during air show. Systems for transmission audiovisual data from fast-moving objects exist (Formula 1 or Red Bull Air Race), there, however, the machines moves over a relatively well-defined route. Therefore, demanding on the radio part is not so high, it is just neccessary to deploy properly the basic stations along the assumed trajectory and the connection is stable and good enough to transfer these data [1]. Nevertheless, this is not possible at aerobatic show because of unpredictable airplane trajectory. Presented solution is designed to increase the robustness of data transmission and reduce the impact of lower transmission conditions by using multiple independent data channels. Fading, that increase error rate, is arise during movement. The proposed methods should reduce the impact of these losses and increase the overall robustness of the system. A common way to improve the transmission characteristics is the MIMO technology (multiple-input multiple-output), thus the use of multiple antennas at the receiver and transmitter. The proposed method is independent of the physical layer and modifies only the second layer of the RM OSI. Therefore, it is universal and it is also possible to use it with e.g. MIMO technology. Looking for optimal parameters, it requires a number of the tests, which is technically and financially very exhausting and demanding, specifically in the case of the aerobatic plane. For this reason, the usage of the simulation is very useful because of an adequately accurate description of the system which allows quick adjustment and evaluation. Therefore, an OMNeT++ was used, the network simulation framework, and library INETMANET that includes support for the international standards that are used in the simulated real device.

II. Implementation

The implementation of the second layer optimization can be done in the different ways. Two goals were chosen for our approach to increase the resilience of transmission (reduce error rate) and consequently to increase the overall the transmission speed. To comply with these requirements, we simulated two new methods and one typical method to compare them. In the selection of the first method, we proceeded from the assumption that if we create N parallel data streams and each will transmits 1/M of input data, then, thanks to a lower bitrate for a single data channel will be achieved higher resistance using more robust modulation. In the second case we counted that the N parallel channels will always transmit the same data. The total bitrate will not increase, but due to parallel transmission of data will be preserved high durability and low error rate. Thus, the resulting bitrate would not be any significantly changed from the nominal speed at physical layer in given modulation.

Figure 1. Block diagram for Case 1

These two approaches were compared with the case using only one radio interface. Its second layer of RM OSI is not modified in any way; the data on the transmitter side comes directly to a single radio module that provides data transmission of one radio channel. The receiving part is again only a single radio interface, and no packet selection is applied. This network was created because of the possibility of comparing the resulting statistics from the other two cases, because this case corresponds to a commonly used method of data transfer - with a one radio interface. This case simulates the situation of the basic wireless data transmission, as indicated by the Figure 1.

Second and third case has functionality differentiated by a distinct block that controls distribution of data for each

interface. This block operates on second layer of RM OSI, is independent on the particular transmission technology or the physical layer. Second case contains the control block on the transmitter side, provides an exact alternate use of both radio interfaces. The principle is shown in the Figure 2. Output gates of block are tied to their own separate radio interfaces transmitting at a different radio channel. The data are collected alternately from the first and the second channel on the receiving side. The simulated network should do this matter, because the load is divided into two channels and provide better error rate compare to a first case, because is viable to use more robust modulation and thus ensure higher availability. It should also lead to increased availability since that there is different interference on each radio channel. Distribution of packets on each channel shows the following set expression witch operate with packets indexes:

$$\begin{aligned} DataIndex &= CH1 = CH2 \\ CH1 &= \{nr | nr = 2k, k \in \mathbf{N}^+\} \\ CH2 &= \{nr | nr = 2k+1, k \in \mathbf{N}^+\} \end{aligned} \tag{1}$$

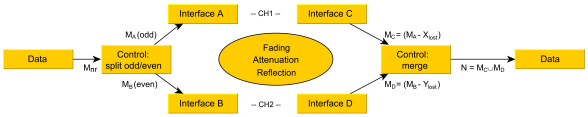


Figure 2. Block diagram for Case 2

The third one, with mirroring of sending messages, generates the duplicate packets in the control block on the transmitter. The original packet and the exact copies have different radio channel again, but this time, these packets are always sent in one moment. On the receiving side, there will be selected the only one of them. The selection is performed at the receiver side in a way to avoid the duplication of the same data. This case is schematically illustrated in the Figure 3.

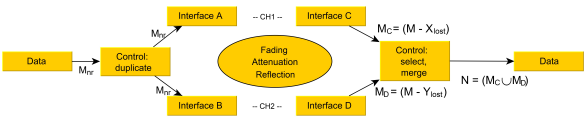


Figure 3. Block diagram for Case 3

Due to high redundancy, the increased availability should results also in this case. The proposal was made for two radio interfaces, but it can be used for a greater number of independent interfaces. This network should show the best resistance to loss of data in the given modulation. Because if a packet will be dropped on one interface (due to faulty transmission, packet damage), there is still the possibility that on the second interface, working on another channel, the error may not occur and therefore the data will be delivered. In fact, it does not matter from which interface the data arrived.

Usage of channels describes this set expression:

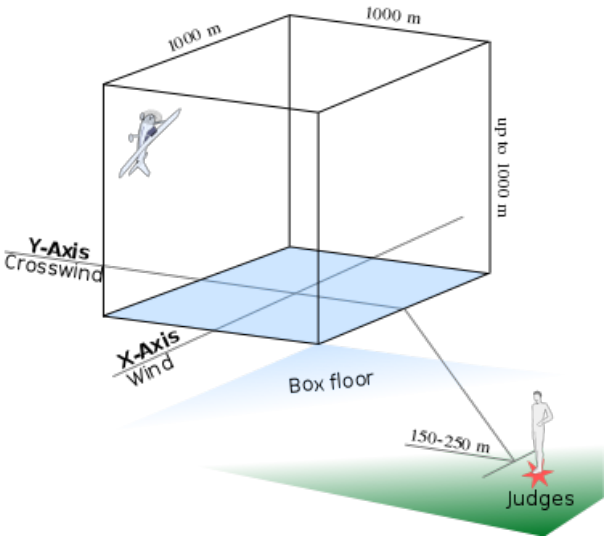


Figure 4. Aerobatic box

$$\begin{aligned} DataIndex &= CH1 = CH2 \\ CH1 &= \{nr | nr \in \mathbf{N}^+\} \\ CH2 &= \{nr | nr \in \mathbf{N}^+\} \end{aligned} \tag{2}$$

Use of the free space model is possible because of precisely defined conditions, where the airplane can move. This is so-called aerobatic box (shown in Figure 4), in which could be always assumed direct visibility and space with no terrain obstacles, vegetation or other objects, so is no need to solve problems by shading the first Fresnel zone. The right direction of antennas was assumed in the simulation. Aerobatics is conducted only in good weather, so the occurrence of rain, or fog is not to be expected. In simulation was calculated with zero gains to the antennas, zero losses caused by unadapted polarization as well as attenuation of the cables and connectors. The same model of propagation is used in all three cases:

$$FSL = 10 \log \left(\frac{4\pi d}{\lambda} \right)^2 \tag{3}$$

is free space loss [-], following (λ - wavelength [m] and d - antennas distance [m]).

From these equations, it is possible to derive the formula for calculating of the signal to noise ratio, which has impact for the error rate.

$$SNR = 10 \log \left(\frac{\frac{P_t G_t G_r \lambda^2}{(4\pi d)^2}}{P_n} \right) \tag{4}$$

where: P_t is transmitting power [mW], G_t, G_r gain of the antennas [-] and P_n is noise power at the receiver input [mW].

The simulation for all three cases was made with the same parameters. One of the changed parameter was the distance between receiver and transmitter, it was necessary to cover

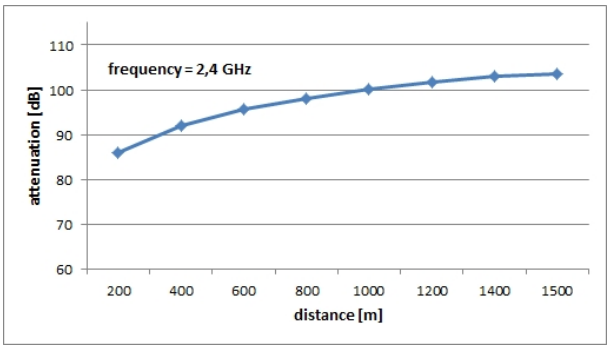


Figure 5. Attenuation of Free Space Loss model

Table I
MODULATIONS AND TRANSMISSION RATES FOR IEEE 802.11G STANDARD

Modulation	Coding rate	Bitrate	Sensitivity
BPSK	1/2	6 Mbps	-82 dBm
BPSK	3/4	9 Mbps	-81 dBm
QPSK	1/2	12 Mbps	-79 dBm
QPSK	3/4	18 Mbps	-77 dBm
16-QAM	1/2	24 Mbps	-74 dBm
16-QAM	3/4	36 Mbps	-70 dBm
64-QAM	2/3	48 Mbps	-66 dBm
64-QAM	3/4	54 Mbps	-65 dBm

the situation where the aircraft occurs in different places in aerobatic box. The selection of distances (200 m, 400 m, 600 m, 800 m, 1000 m and 1500 m) is given by the conditions in which the aerobatic planes are moving. These conditions are set by the contest rules, as shown in Figure 4. The maximum distance has a value of 1500 m, which roughly corresponds to the largest distance achievable in a given area. The lower limit is again determined by the conditions in which the planes moves normally and is around 200 m above ground. The progression of attenuation in these distances is shown in Figure 5.

The next changed parameter was transmitter power, its value was in range from 20 mW to 1000 mW (20 mW, 50 mW, 100 mW, 300 mW, 500 mW, 800 mW a 1000 mW). Selected values was given by the binding simulation with already realized project. In particular, values below 10 mW was insufficient even for the realization of the transmission on the lowest distance in combination with the most robust modulation. The last changed parameter was transmission rate, values were taken from IEEE 802.11g standard [2] shown in Table I. This standard was chosen due to its common availability and ease of use.

The appropriate modulation and coding rate according to the selected transmission speed automatically provides INETMANET library in compliance with the selected standard. This standard defines the minimal value of sensitivity and code rate, which indicates the ratio of the number of bits of useful information to the total number of bits (including the added redundancy), for each type of modulation. Selected parameters (for channel bandwidth 20 MHz) are in Table I. Simulations were performed for all permutations of the above parameters for each of the proposed case. In total, there was evaluated data from more than 1000 simulations.

III. SIMULATION FRAMEWORK

To simulate the transmission protocols and technologies, there are the most suitable discrete simulation tools, as directly derived from the nature of the analyzed and simulated problem. One of the representatives of the discrete event systems is OMNeT++, which was used to create a model representing the transmission system. OMNeT++ is object oriented, modular system, based on the processing of discrete events. Possibilities of its use include a wide range of areas, it can be used to create simulations of any system that can be described using discrete events, which can be decomposed into an elements communicating together with messages [4]. There are a number of preconfigured models for it, standards and technologies in libraries targeted to a specific area. Due to the implementation of common standards is feasible to reduce number of simplifications in the model. For simulations of data transmission from the aerobatic planes was used INEMANET library, which contains models for several wired and wireless networking protocols, including UDP, TCP, IP, Ethernet, PPP, 802.11, and many others [3]. Part of the library is also implementation of the behavior of the physical layer, such as different models of signal propagation, modulation type, sensitivity, or transmitter power.

IV. ANALYSIS

The simulation of error rate in network framework is provided by a set of experimentally obtained data from INETMANET library. The set of data contains a lines on which is defined SNR (in dB units) and the probability of loss for a certain bit rate and packet size [6]. Some values obtained in this study, paradoxically, with greater distances descends. This is due to the set of data, which models the probability of packet error. Its function is based on comparison of the size (in Bytes) and SNR of packet that arrives on the radio interface, with the line of that set of data on which is defined loss rate of the packet for a given speed (modulation type). Values in this set was obtained experimentally at University of Padova (others can be one interpolated) [6]. In the case of a match, is the packet considered to be erroneous with the set probability.

Value of PER (packet error rate) raising with increasing bitrate (related to the used modulation and code ratio). Sufficient value of packet loss for transmitting high-speed video signal with H.264 compression (10% according to [7]) achieved all cases at bitrate up to 12 Mbps for all required distances. It is possible to operate at 24 Mbps with all 3 cases at distances up to 800 m. The Figure 6 shows the growth of error rate with increasing distance between receiver and transmitter for transmitting power of 100 mW. This chart intentionally shows distances greater than the rules of aerobatic contests required due to demonstrate behavior of cases two and three to the

first case. Case 2 shows more robust modulation BPSK with code rate of 3/4, while the other two cases are in the QPSK modulation with coding rate 3/4. Case 1 remains below the desired limit of error rate up to the distance 1200 m, at 1400 m is connection already broken. Case 3 ends connection at this distance as well, but it is able to maintain a significantly lower error rate up to this value compared to the Case 1. Case 2 benefits from lower and thus more robust modulation, which is more reliable, therefore is capable sustain the required error rate limit up to a distance of 2100 m. This method shows, in the same range of distances, similar or better characteristics to a first case. Therefore, is confirmed that for the given modulation scheme is the Case 3 the most suitable because of keeping a very good error rate. Usage of the Case 2 is very advantageous to keep the error rate below a certain threshold even for longer distances.

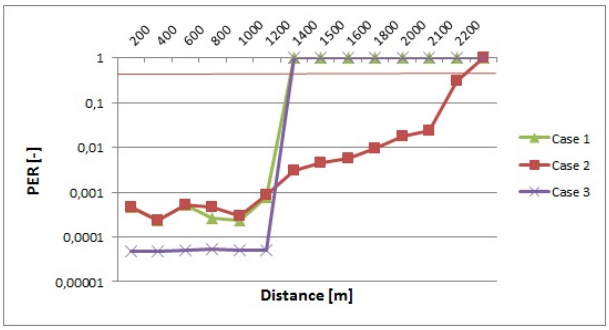


Figure 6. PER for all 3 cases at 100 mW of transmitting power

V. Conclusion and Future Work

Simulation of transmission system of the aerobatic aircraft was created by network simulator OMNeT++ and its library INETMANET. Two types of modified methods of sending data were compared with usual wireless transmission on this simulation model. The modified transmission is based on the usage of the multiple independent radio channels and it is carried out at the second layer of the OSI model; is not dependent on the particular transmission technology, physical layer is not affected as well. This option increases the versatility of the solution. The case using two radio interfaces on both sides of the transmission system sends the same data from both interfaces, but on different channels. This method proved to be very robust, because the receiver, in case of a damaged packet from the first channel, automatically receives data on the second channel. This way achieved the lowest error rate for a given modulation. The case using alternating sending packets on the two interfaces on different radio channels confirmed the benefits of maintaining a higher overall data rates for longer distances, thanks to the use of stronger modulation in both channels.

The future work will aim at improving the proposed design for more accurate description of real situation. That involves implementation of movement in the 3D scene with usage of data that was measured during aerobatic competition. Also there is a possibility to use more precise description of communication system, e.g. specify antennas characteristic or polarization mismatch and utilize the GPS (Global Positioning System) data to connect the model with real position above surface in a map. In addition to the simulation video broadcasts of the fast moving objects, we would like to incorporate the possibility of the transmitting voice data into the model, focusing on the SIP (Session Initiation Protocol) network infrastructure. The issue of real testing of SIP is very well described in the [8] and [9]. New simulation model, which will be incorporated into the new version of the developed simulation library, is now created according to the information and procedures described in these articles.

Acknowledgment

This work was supported by grants MSM6840770038 and SGS11/124/OHK3/2T/13.

References

[1] Machula, V., Kocur, Z. Možné způsoby přenosu audiovizuálních dat z akrobatických letadel. Access server [online], [accessed on March 2011], http://access.feld.cvut.cz/view.php?cisloclanku=2010120001

[2] IEEE Standard for Information Technology-Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, IEEE Std 802.11-2007, http://standards.ieee.org/getieee802/download/802.11-2007.pdf

[3] INET framework home page, [accessed on November 2010], http://inet.omnetpp.org/

[4] Varga, A. *OMNeT++ User Manual*, Version 4.0, 2010, http://www.omnetpp.org/doc/omnetpp40/Manual.pdf

[5] Chlumský, P. *Optimalizace chování bezdrátových datových protokolů pro mobilní aplikace*, Diploma Thesis, 2010

[6] OMNeT++ Discrete Event Simulation System – community site [online], [accessed on April 2010], http://www.omnetpp.org/

[7] Fallah, Y.P.; Koskinen, D.; Shahabi, A.; Karim, F.; Nasiopoulos, P.; A cross layer optimization mechanism to improve h.264 video transmission over wlans. In Consumer Communications and Networking Conference, 2007.CCNC 2007. 4th IEEE, 2007.

[8] Voznak, M., Rozhon, J. SIP Infrastructure Performance Testing, p. 153-158, In Proceedings of the 9th International Conference on TELECOMMUNICATIONS and INFORMATICS , Catania, Italy, May 29-31, 2010, ISBN 978-954-92600-2-1, ISSN 1790-5117, May 29-31, 2010

[9] Voznak, M., Rozhon, J. Methodology for SIP Infrastructure Performance Testing , WSEAS TRANSACTIONS on COMPUTERS, Issue 11, Volume 9, pp. 1012-1021, September 2010, ISSN: 1109-2750.

ZigBee Sensor Network for controlling the lightening system

Bohdan Borowik
Dept. of Electrical Engineering and Automation
University of Bielsko-Biala
Bielsko-Biala, Poland
bo@borowik.info

Barbara Borowik
Dept. of Physics, Mathematics and Computer Science
Cracow University of Technology
Cracow, Poland
cnborowi@cyf-kr.edu.pl

Igor P. Kurytnik
Head of Dept. of Electrical Engineering and Automation
University of Bielsko-Biala
Bielsko-Biala, Poland
ikurytnik@ath.bielsko.pl

Volodymyr Karpinskyi
Dept. of Electrical Eng. Power Consumption Systems
Ternopil Ivan Pul'uj National Technical University
Ternopil, Ukraine
vkarpinskyi@gmail.com

***Abstract*— A control system for an HID (High Intensity Discharge) lamp with the magnetic ballast is considered. Electronic switch unit driven by ZigBee system controls illumination of lamps in network. RF ZigBee transceivers are employed to turn the HID lamps on and off and to control illumination trough wireless remote communications. Thereby are achieved considerable energy saving. Also operational status of the balaststst and HID lamp can be controlled by the sensors and all failures will be automatically detected and notified to users.**

Keywords-HID lamp; ballast; Zigbee network, PIC microcontroller; Xbee module

I. INTRODUCTION

The subject of the paper is our investigation with ZigBee module (802.15.4 protocol) for controlling HID lamp system. It is scalable and prepared for municipal and high-way applications. Schematic diagram of the system is shown in figure 1.

Because a certain area receives a plenty of sunlight in daytime and does not need lighting in that time, whereas another areas do need, illumination of lamp should be controlled, in order to save energy. The magnetic ballast control system controls illuminations and easily grasps operation and failure status of the lamp (through current sensor). The control system for a HID lamp can have a low-priced magnetic ballast, an electronic switch unit for controlling Illumination . RF transmitting and receiving modules, that use the ZigBee based frequency, are deployed, so that turning on and off and illumination of the HID lamp can be remotely controlled.

An HID (high intensity discharge) lamp is a lamp, in which high pressure xenon gas and metallic compounds in arc tube are excitedly discharged and generate white light close to natural, since the xenon gas suppresses neighboring ultravionet rays and electromagnetic waves. The HID lamp, such as a metal halide lamp, high pressure sodium lamp, high pressure mercury lamp and the like, is lit by a specially designed ballast (magnetic or electronic) and a starter that generates lamp operation .

Each lamp has its own wireless control module with Zigbee module, CPU, application board with sensors, power supply, electronic switch unit and ballast unit. Dispatcher of all lamps sends from his office wirelessly signals to particular lamps and receives feedback. In addition, the module can be controlled through a gateway and a Zigbee wireless network.

The Zigbee technique [6] is a standard specification based on IEEE 802.15.4. It has a low transmission speed and can remotely control a variety of equipment. The Zigbee PHY layer uses a frequency bandwidth of 2.4 GHz and 868/915 MHz, direct sequence spread Spectrum (DSSS) modem and MAC of CSMA/CA. With a communication distance of 100 to 500 meters it consumes less power, than 500 μW.

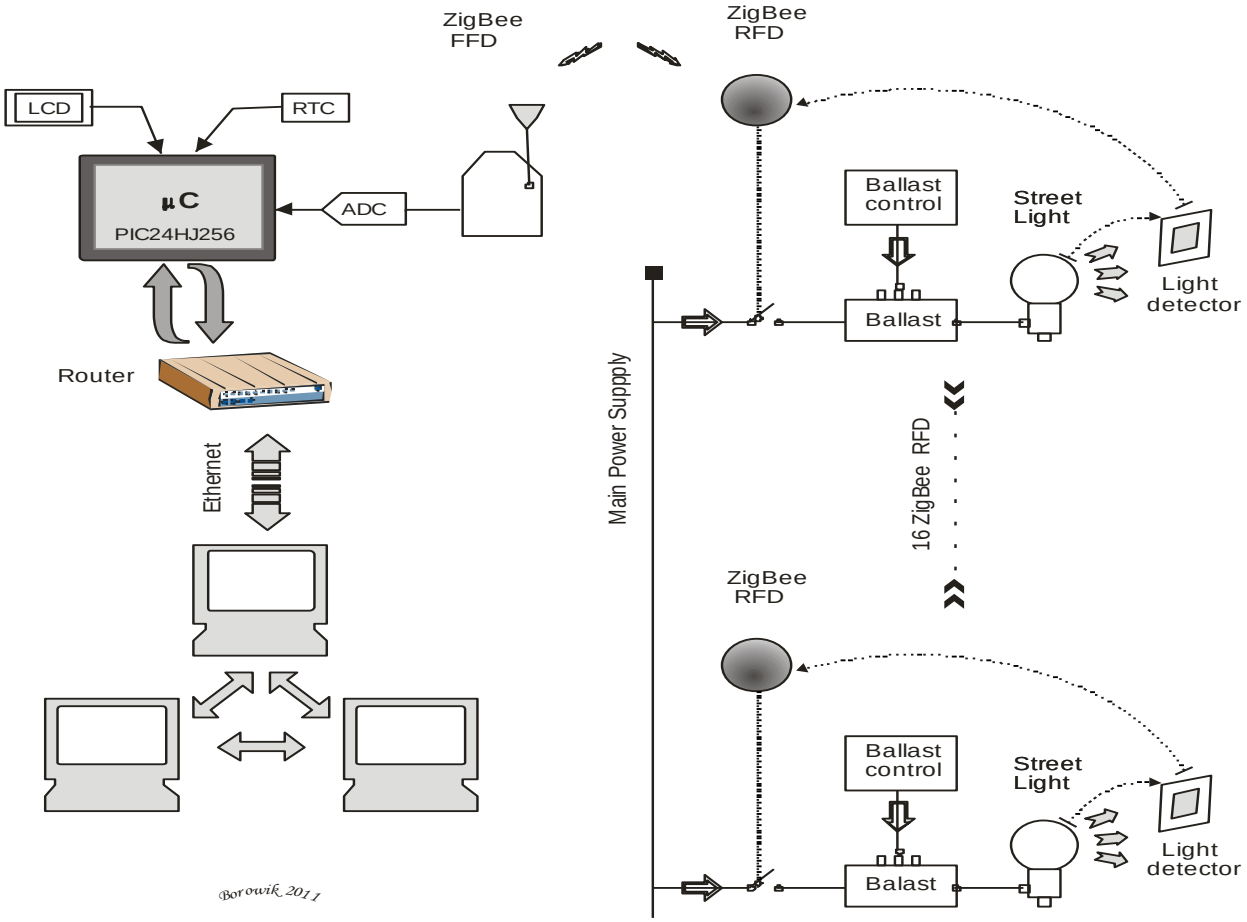


Figure 1. Schematic diagram of the wireless control module system

II. DESCRIPTION OF ZIGBEE MODULE

In this paper we propose some detailed solution of the system. As CPU in the lamp control module the Microchip PIC 24HJ256GP610 microcontroller [7] is used. It has following tasks:

- acquiring data from the current detecting sensor of the HID lamp, thus detecting, if the lamp is not broken,
- receiving signals from dispatcher or from internet through gateway and the Zigbee wireless network,
- processing received signals
- dimming lamp lightening and turning lamp on and off through generating appropriate signals to electronic switch unit and subsequently to the magnetic ballast unit (see figure 2.)
- sending feedback to the remote management module through Zigbee RF transceiver.

The PIC 24HJ256GP610 microcontroller is well suited to acquiring analog data from sensor, because is equipped in several inherent ADC modules. We used the ADC module with the 10-bit resolution. After selecting port line for analog input we prepared in C language [10] the functions for initializing and configuring ADC module and for acquiring and converting data from the analog input [11]:

```
// initialize the ADC for single conversion, select Analog input pins
void initADC( int amask)
{
   AD1PCFGH = 0xFFFF;
         AD1PCFGL = amask;    // select analog input pins
   AD1CON1 = 0x00E0;   // auto convert after end of sampling
   AD1CHS0  = 1;        // select analog input channel
   AD1CSSH = 0;       // no scanning required
    AD1CSSL = 0;
   AD1CON3 = 0x1F02;   // max sample time = 31Tad, Tad = 2 x Tcy = 125ns >75ns
   AD1CON2 = 0;       // use MUXA, AVss and AVdd are used as Vref+/-
   AD1CON1bits.ADON = 1; // turn on the ADC
    } //initADC


//--------------------------------------
unsigned long readADC( int ch)
{
    AD1CON1bits.SAMP = 1;       // start sampling, automatic conversion will follow
```

```
        while (!AD1CON1bits.DONE);   // wait to
complete  the conversion
        return (long) ADC1BUF0;   // read the conversion
result
    } // readADC
```

If the changes in the current are substantive, the failure of the magnetic ballast or the HID lamp is assumed.

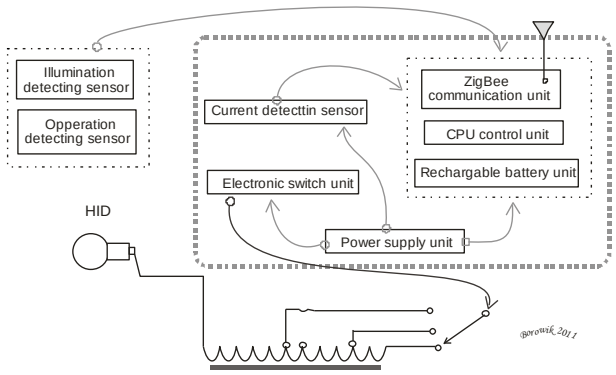


Figure 2. Solution for the magnetic ballast driver

As Zigbee units we used the DIGI International Xbee modules [8]. They can operate in the API programmed mode, which is convenient in our case. It is the frames mode. Frames of selected type, containing payload with data are sent from microcontroller to UART. They are send to remote XBee modules. Payload can contain measurement results from sensors as well as feedback from dispatcher. Used transmitted frame types are shown in the Table I.

TABLE I. API FRAME TYPES.

API ID	API Frame Names
0x10	ZigBee Transmit Request
0x17	Remote Command Request
0x8B	ZigBee Transmit Status
0x90	ZigBee Receive Packet (AO=0)
0x97	Remote Command Response

Illumination is controlled by setting one of four values for the light outputted: 100%, 80%, 70% and 50%. It is accomplished in the electronic switch unit by providing condensers directly connected to the HID lamp. When condensers 10, 3 and 2 μF are used, the following output values can be obtained:

TABLE II. LIGHTENING OUTPUT FOR DIFFERENT CONDENSERS

Operation	10 μF	3 μF	2 μF	Total capa-citance
Turm on Output 100 %	Used	Used	Used	15 μF
Output 80 %	Used	Used	Unused	13 μF
Output 70 %	Used	Unused	Used	12 μF
Output 50 %	Used	Unused	Unused	10 μF
Turn of	Unused	Unused	Unused	0

The dispatcher decides what dimming to apply to the lamp and sends appropriate information to the wireless control module. The information is received by the PIC microcontroller from remote in the interrupt service routine and the value for the output flag is set. XBee module, working as ZigBee transceiver has also 11 analog and digital I/O lines. Xbee module can have up to 11 analog and digital IO lines. Pins 17 thru 20 can be configured as analog input lines.

Table III. Analog Input pins

pin number	pin name
17	AD3/DIO3
18	AD2/DIO2
19	AD1/DIO1
20	AD0/DIO0/ Commissioning Button

To enable an analog function on XBee module pin(s), the appropriate configuration command must be issued with the correct parameter: parameter 2 enables analog input.

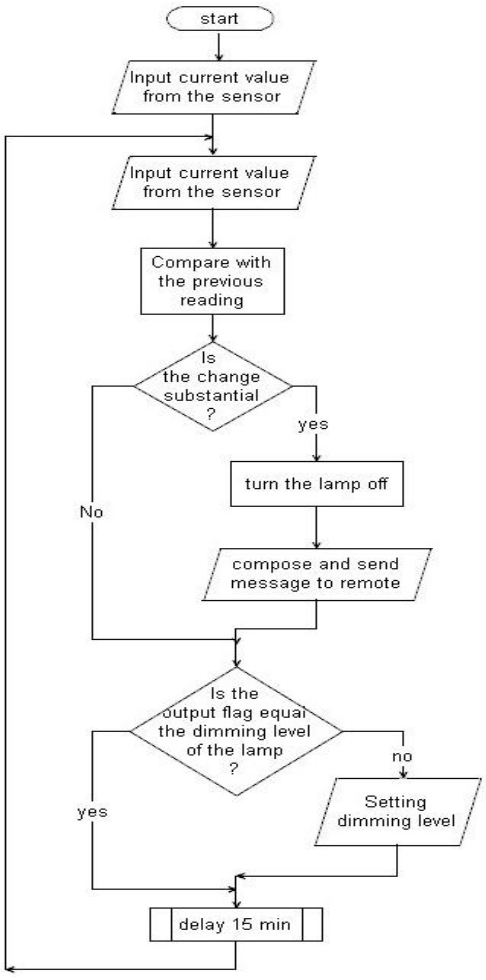


Figure .3. Flowchart of the firmware permanent loop

Then issuing remotely IS command forces a read of enabled analog input lines. XBee has also command allowing to obtain I/O samples periodically. I/O sample can bee taken and transmitted to a remote device at a periodic rate. The command is IR. It requires parameters specifying periodic rate. Parameter range is 0 – 0xFFFF ms. The value 0xEA60 (Dec 60 000) sets periodic rate to about 1 min (60 sek). After configuring and initializing XBee module and the microcontroller, the permanent loop is executed on the basis of the time interval 1 min. performing all necessary actions:

- checking the current value from the sensor and comparing with the previous reading
- if substantial changes in the reading are detected, the signal is generated, that cuts off power of the failed lamp
- checking the value of the output flag and comparing with the dimming level of the HID lamp. If the values differ, setting the new value to the lamp lightening through generating signal to the electronic switch unit.

The above actions can be schematically presented on the diagram (Figure 3).

III. DISCUSSION

As we described in our previous work [8], the real time throughput in ZigBee network is relatively slow and well suited for low frequency sampling. In such a case all system operations are stable.

XBee module has also mechanism of retransmitting when packets are delivered with errors. Our investigations were enhanced with the ZENA Wireless Network Analyzer. It is equipped with hardware Packet "SNIFFER" and can graphically display decoded packets, as well, as network topology. From ZENA Packets Sniffer window it became possible to learn all composition of the traffic in the selected channel: polling requests from ZED (ZigBee End Device), ACK (Acknowledge) frames from ZC (ZigBee Coordinator), frames with data and frames with ACK from ZC generated from different layers. Parent node in Zigbee PAN network is coordinator. If coordinator module has on its board microcontroller and Mass Storage Device, then additionally it is possible collecting results from sensors on mass storage device such as SD Card, for further analysis and processing.

IV. CONCLUSION

In many industrial processes, one of the major problems is the data sharing by cable connections. Producers of the ZigBee wireless network devices equip them with the set of interfaces enabling a connection of miscellaneous measurement sensors and converters. It is possible to select the adequate converter allowing of the full utilization of capability of the ZigBee network.

Among several producers, whose devices and testing boards were used in our investigation, Digi's Xbee modules appeared to be adequate and fully relevant in sensor data monitoring.

Our investigation lets presume the feasibility of deploying the wireless sensor network for HID lamp control system, by using the PIC24HJJ256 microcontroller and XBee transceiver modules.

REFERENCES

[1] I.P.Kurytnik, M. Karpiński "Bezprzewodowa Transmisja Informacji", Wydawnictwo PAK Warszawa 2008.

[2] B.Borowik I.P. Kurytnik "Mikrokontrolery PIC w zastosowaniach", Wydawnictwo PAK, Warszawa 2009.

[3] B. Borowik "Tilt and vibration mesurement of the remote objects using ZigBee communication" // Przegląd Elektrotechniczny.– 2010. – Nr 7.– S. 57-59.

[4] B.Borowik "Wykorzystanie sieci sensorycznej ZigBee do monitoringu układów dynamicznych", BEL-Fund 2009.

[5] Borowik B. New approach to measuring vibration parameters of the remote objects with the ZigBee technique //16th IMEKO TC4 Symposium Exploring New Frontiers of Instrumentation and Methods for Electrical and Electronic Measurements Sept. 22-24, 2008, Florence, Italy

[6] www.zigbee.org

[7] PIC24HJXXXGPX06A/X08A/ X10A High-Performance 16-bit Microcontrollers Data Sheet, 2009 Microchip Technology Inc., DS70592B.

[8] Xbee®/Xbee-PRO® ZB RF Modules, Digi International, 2010.

[9] B. Borowik, J. Wojnarowski, „Applying the Zigbee technology for the enhancing the remote objects control", 80 Annual Meeting International Association of Applied Mathematics and Mechanics (GAMM2009) Gdansk University of Technology, February 9-13, 2009.

[10] B. Borowik et al., Meandry języka C/C++, PWN, Warszawa 2006.

[11] B. Borowik, "Interfacing PIC Microcontrollers to Peripheral Devices," Springer Science + Business Media B.V. 2011.

[12] B. Borowik, "Wireless sensor network for monitoronig vibration of dynamic objects" BEL-Fund. Kraków,2010.

DTMF signal Generated by Processor PIC18F452

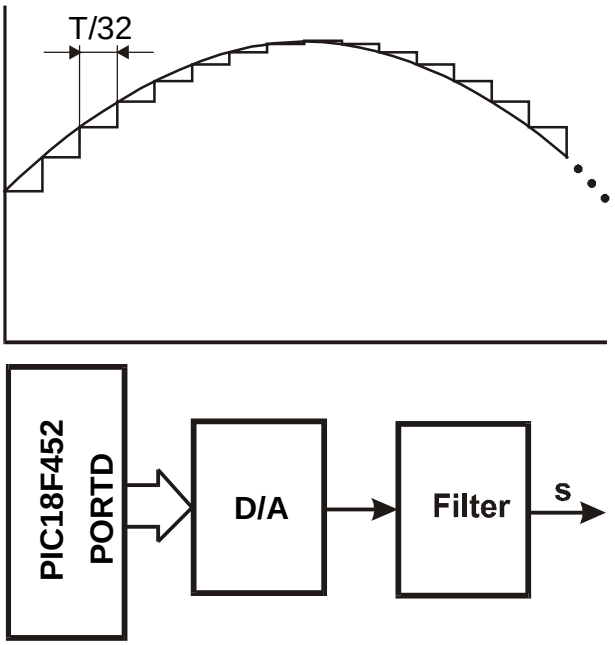
Radek Novak
Department of Telecommunications
VŠB TU Ostrava, FEI
Ostrava, Czech republic
radek.novak@vsb.cz

***Abstract*—An application of PIC18F452 uses timers Timer1 and Timer3 for generating DTMF signal. Output signal is filered by Cebysev low-pass filter. Also simple keyboard 4x3 is connected.**

Keywords- DTMF signal; PIC18F452; Cebysev low-pass filter

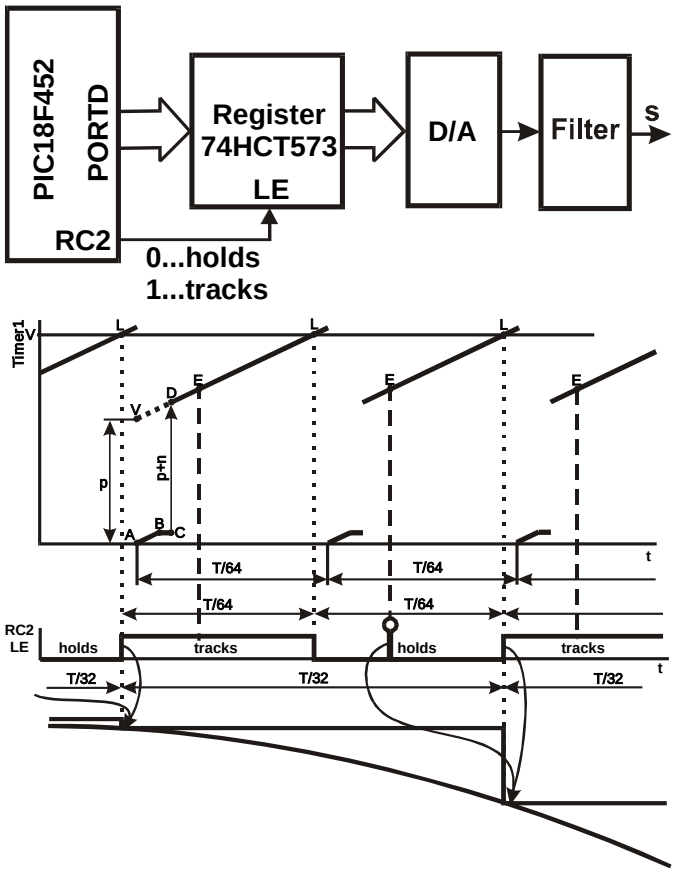
I. INTRODUCTION

Processor generates two internal sinus signals marked as "a" and "b". Their frequencies are f_a(697,770,852,941Hz) and f_b(1209,1336,1447,1633Hz). After some experiments was decided to divide full period T of sinus into 32 intervals T/32(concretely T_a/32 T_b/32) and generate sinus as "stair signal". Application uses crystal 9,316MHz and internal PLL, so final frequency is 36,864MHz and duration of machine cycle is 108,507ns. Basic of all works was made with simple sinus generator realized by PIC18F452 and surrounding circuits. Block schema of it is on the pic. 1.



Pic. 1 Sinus approximated by stair signal and basic sinus generator and block schema of the simple generator

Simple solution from pic. 1 had some inaccuracy. Description of this inaccuracy - time increment T/32 is measured by internal timer in processor PIC18F452. If this timer overflows on instruction taking 1 machine cyclus, then generating of stair signal is absolut accurate. But few instructions of PIC18F452 take 2 machine cycles, program execution enters into interrupt service routine for 1 machine cyclus later, and also sending new value to PORTD (for D/A converter) is 1 machine cyclus later. Using an oscilloscope we can see some jitter of transition edges of stair signal. This problem was solved by using latch register 74HCT573 and dividing increment T/32 into two half intervals T/64 + T/64, see pic. 2.



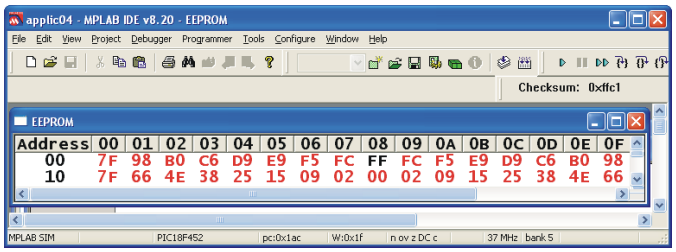
Pic. 2 Exact solution for generating sinus using register

Exact solution has this principle. Timer1 now overflows 2x more frequently i.e. with period T/64. Solution uses Compare facilities of Timer1 . An overflowing of Timer1 does not invoke interrupt, but Compare event yes. Actual value of Timer1 is by compare logic all the time compared with value "V" (16 bit). Suit value of "V" must be big – for example FFFE. In case of equality (point L) an interrupt request is generated. Compare logic also can total automatically

invert(toggle) output pin RC2, that's an excellent facility for this application. Running cyclus of Timer1 in itself is periodical exactly. Near before overflowing compare logic detects equality, an interrupt is invoked, and pin RC2 is automatically inverted. Meantime the Timer1 overflows and falls to point A. Than the Timer1 will be stopped by instruction, so it gets point B. Timer1 is incremented by instruction about value p+n, it transits into point D. Subsequently (point E) is necessary to read actual value of LE signal(RC2 pin). If read value is 0, so register holds, and it is suit moment to prepare new value onto inputs of this register(new valued is contained in internal Data EEPROM). Moment of preparing this new value is in pic.2 marked by a ring. New value will appear on outputs of register at LE=1, it will come at next equality Timer1 with value "V". Position of point "E" can little vary (reason is the same as in solution without the register), but it has no influence to time duration of stair elements, thanks to using register.

II. SOLUTION OF DTMF GENERATOR

Having experiences from creating exact sinus generator, then the DTMF generator was realized. The first realization used two ports of PIC18F452, and signals "a" with "b" was added outside of processor – using operating amplifier. It was temporary solution and will not be descripted in this paper. The second(final) realization consists in one sinus table(in Data EEPROM) and using one PORTC only. 32 values of sinus are calculated and saved into Data EEPROM in introduction part of program :

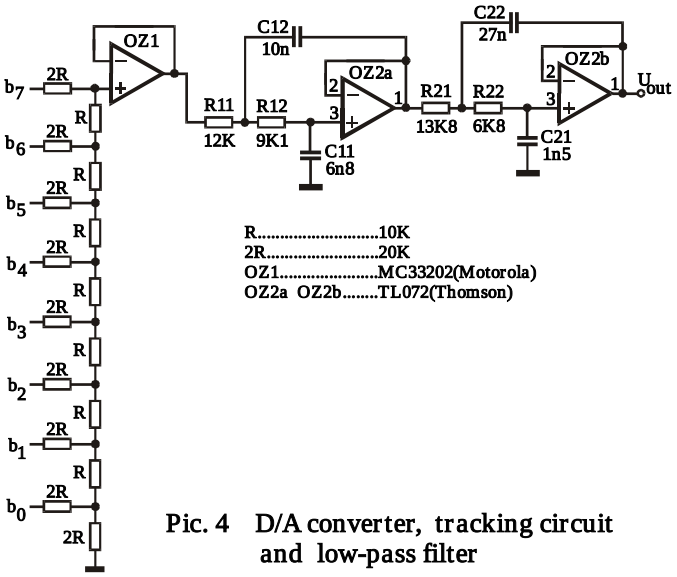


Pic. 3 Sinus table contained in Data EEPROM

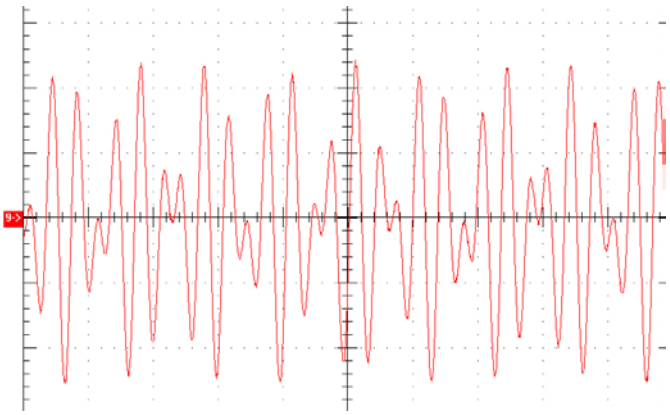
For orientation in table serves two pointers pointer_a pointer_b, their values are modulo 32. The most fast method (rather for PIC18F452) how to limit pointer into range 0-31 is clearing bit b5 after every incremenation , see next two lines from program :

```
++pointer_a;
pointer_a &= 0b11011111;
```

For measuring time intervals $T_a/32$, $T_b/32$ are used timers T1, T3. D/A converter (for the reason maximal speed) was realised by resistance ladder, connected to 8-bit PORTC. Resistance ladder is followed by tracking circuit, based on OZ1. There is a Cebysev low-pass filter connected in the end of all the circuit, based on OZ2.



Pic. 4 D/A converter, tracking circuit and low-pass filter



Pic. 5 Final DTMF signal at pressing button 7 (f_a=852Hz f_b=1209Hz). Measured by Tektronix TDS2014.

CONCLUSION

Obtained results were excellent, coupled DTMF decoder (type HT9170) had no problems with decoding of this signal. Some little inaccuracy in program generating sinus curves "a" and "b" consist in common interrupt service routine for both timers – Timer1 and Timer3. But this little inaccuracy is practically total hidden by consequential low-pass filter.

Radek Novák *is an assistant in VŠB Technical Univerzity Ostrava, Department of Telecommunications*

REFERENCES

[1] R. Novak, V. Vasinek, "System for Evaluating Impulse Signals," Conference KTTO 2010, Ostrava, December 2010.

[2] R. Novak, "New Application of GSM Module SIM300D", Conference RTT 2010, Velke Losiny, September 2010.

Design of CLI for VoIP routers and gateways

Lukas Macura

Silesian University In Opava
Faculty Of Bussines Administration In Karvina
Email: macura@opf.slu.cz

***Abstract*—This paper deals with configuration management of VoIP systems. Many VoIP system today has its own config files. Some software uses text config files, other uses SQL db and there are some kinds of software which uses binary files. Configuration is possible using text files and web GUI. Bigest problem today is how to represent unified config for all systems. We want to define some higher level language or script, which will define some basic rules for all systems. It means, when we export config from one system, we can use it on another system. Another point of view is simplicity. Administrators would learn only one configuration syntax or cli and software developers will adopt their systems to understand this syntax. After discussion, we would want to write RFC, which will define syntax of this config.**

***Index Terms*—VoIP, SIP, H.323, IAX, ISDN, Asterisk, FreePBX, OpenScapeVoice, Embedding, Phonyx, RFC**

I. Introduction

New config system should be robust, easy and clear to read. It will be hierarchical config with more contexts. Variables and features should be inherited from upper level tree to lower. For example global parameter for entire system should be inherited into all subtrees until not changed. Entire system explained in graph. It is important, that configuration should be as clear as possible. It is better to use entire words, not just abbreviations. Any administrator with good level of knowledge should be able to read and write config parts without big further studying of language.

II. State of the Art

Today, there are some RFCs which focus on similar problematics. Most used is probably NETCONF (RFC 4741), next asynchronous event notifications (RFC 5277), partial locking of running configuration (RFC 5717), YANG (RFC 6020) and common YANG data types (RFC 6021). Mission of this paper is to use power of all this RFC, not to make some concurrence to them. All NETCONF RFCs are focused mainly to generic network devices. For VoIP router and gateways, some specific parameters will be needed.

III. Semantics

Entire configuration system is hierarchical (see IV). There are three possible semantics of configuration syntax. One is *cli* syntax (like shell syntax with autocompletion possibility), second *machine* syntax and third is *selective* syntax.

A. CLI syntax

This kind of configuration is practical for administrator to configure system interactive using commandline interface. Parameters are entered inside containers, in selective contexts. Contexts are changing automaticaly, when right keyword for context is used. See listing 1. If system is configured by this syntax, there can be situations, when configuration for some part of config is not done yet. Typical example is, that we are creating new engine and we have to fill at least one protocol. At this time, configuration is inconsistent until all expected lines and values are entered. Transaction automaticaly starts when entering context and ends when closing context. When configuration is not consistent for context, it cannot be commited. In this situation, we have to either add missing lines or revert transaction. By entering end command, we apply changes for context and return to upper level.

Listing 1. CLI syntax example

```
(.)# container acme.org
(acme.org)# engine kamailio at 10.1.1.1
(acme.org/e-kamailio)# voice-protocol sip
(acme.org/e-kamailio/p-sip)# domain x.y.z
(acme.org/e-kamailio/p-sip)# end
(acme.org/e-kamailio)# end
(acme.org)#
```

B. Machine syntax

Machine syntax is used for importing parts of config or entire config. For easier parsing, groups and containers are enclosed in brackets. It is easier for machine to read well structured text. See listing 2. This syntax will profit from YANG syntax.

Listing 2. Machine syntax example

```
container acme.org {
 engine kamailio at 10.1.1.1 {
  voice-protocol sip {
   domain x.y.z
  }
 }
}
```

C. Selective syntax

This syntax is used, when refering single value of config, data or variable. This can be used even for variable expansion or for reporting informations about system. Each object in config has its own unique identifier. It is possible to get or set value by this syntax. General form of URI is refered on 3.

Listing 3. General form for selective syntax

```
[.context/]@user
[.context/]#extension
[.context/]%cfg
[.context/]$var
[.context/]&map(key)
```

There are this kind of options:

- context/ refers to some context
- @user refers to user in current context
- ċontext refers to context
- #extension refers to extension
- ċontext/#extension refers to extension in directed context
- %cfg refers to config value in current context
- ċontext/%cfg refers to config value in directed context
- $data refers to variable or status value
- &map(key) refers to map value (see map tables)

See listing 4 for examples to set or get parameters from system.

Listing 4. Selective syntax example

```
set .acme.org/%engine/kamailio/10.1.1.1/domain x.y.z
get .acme.org/%engine/kamailio/10.1.1.1/domain
get .acme.org/@director
get .acme.org/#12345
get .acme.org/@director/%email
get .acme.org/@director/$credit
```

D. Variables

Variables can be either static or dynamic. Static variables are part of config and we can refer to them. Their value is known during script parsing. See 4. Dynamic variables are stored on target system and can be changed during time without regenerating config file. For example, call credit can be such information. It is decreased after each call and it has to be saved and managed on target system. Not all engines support dynamic variables, while any engine supports static variables because config for engine is made when variable is known. Variables can be used on any place where text information is expected. It has to start with %, , #, ȯr $ character.

E. Capabilities

Capabilities are not part of config at all. Instead of this, it is hard-driven part of each engine or parser. Similar to *C* language and #define, it is impossible to change capabilities by config. It is possible to see capabilities of engine or map, but it cannot be changed. For example, kamailio knows only SIP protocol. So if somebody wants to use kamailio for H.323, parser should stop and report error. Similar, there can be some external map, where we can search for key, but we cannot list all objects inside. This vapabilities will be mapped to features in YANG.

F. External mapping

For bigger systems, we need to have external databases. For example, we do not want to define all users in config file but we want let system looks into LDAP directory and do its work. Similar to extensions or local variables. There are two kind of external mapping - dynamic and static. Static mapping means that config parser should open map database and generate target systems config staticaly. So it will be actual only during config parse or using explicit command for reload map. Instead of this, dynamic maps are looked up just in time when query occurs. For example, static LDAP map will query LDAP during parsing, get all needed records and generate target config for all users found during query. But dynamic map will make query in start of each call. Some engines are capable to look in some dynamic maps. It highly depends on engine. For example, kamailio or asterisk can look into LDAP for some data, while Cisco voice gateway cannot. See 5. We define map table1, which takes telephonenumber as argument, searches for telephonenumber and returns uid of found user. Maps are more complex, can use variery of DB and are out of scope of this document.

Listing 5. External maps example

```
container acme.org {
 static-map table1 on LDAP with keys tn {
  binddn uid=user,dc=acme,dc=org
  bindpw password123
  root dc=acme,dc=org
  search (telephonenumber=$tn)
  return uid
 }
 extensions at table1
}
```

IV. Container model

See figure 1 for overall lookup of containers and hierarchy. Container is group of setting which are glued together and are hierarchical. It is very similar to domain names and in fact, we should use real domain names for containers, where possible. For example, company acme.cz should use left part of tree and acme.eu right part of tree. Design is even similar to LDAP tree. And in fact, there should be no problem to save config for some part of tree into LDAP using external maps. Parameters are inherited from upper level container into lower level. Users can be found in any part of tree, similar to LDAP. Basic principle of design is, that we can at any time cut some part of tree and make its own config for it. For example, if acme.cz will fork from global ACME, config of their tree can be used for their system and will be external to acme.eu. It is important to understand this principle. Configuration can be as simple as one context for one PBX system in one company or more complex for many PBX systems of more companies.

A. Engine

Engine means software or hardware which works as voice switch. It can be variety of software or hardware. It can act as primary or secondary for container. It is possible for one engine to serve more containers. Primary egine is responsible for call routing in normal situation. Secondary engines are used to serve when primary fails. This configuration does not try to describe high availability systems with shared IP like VRRP. From our point of view, entire HA system is one engine. Our config system uses secondary engines only for protocols which support it and for clients which understand

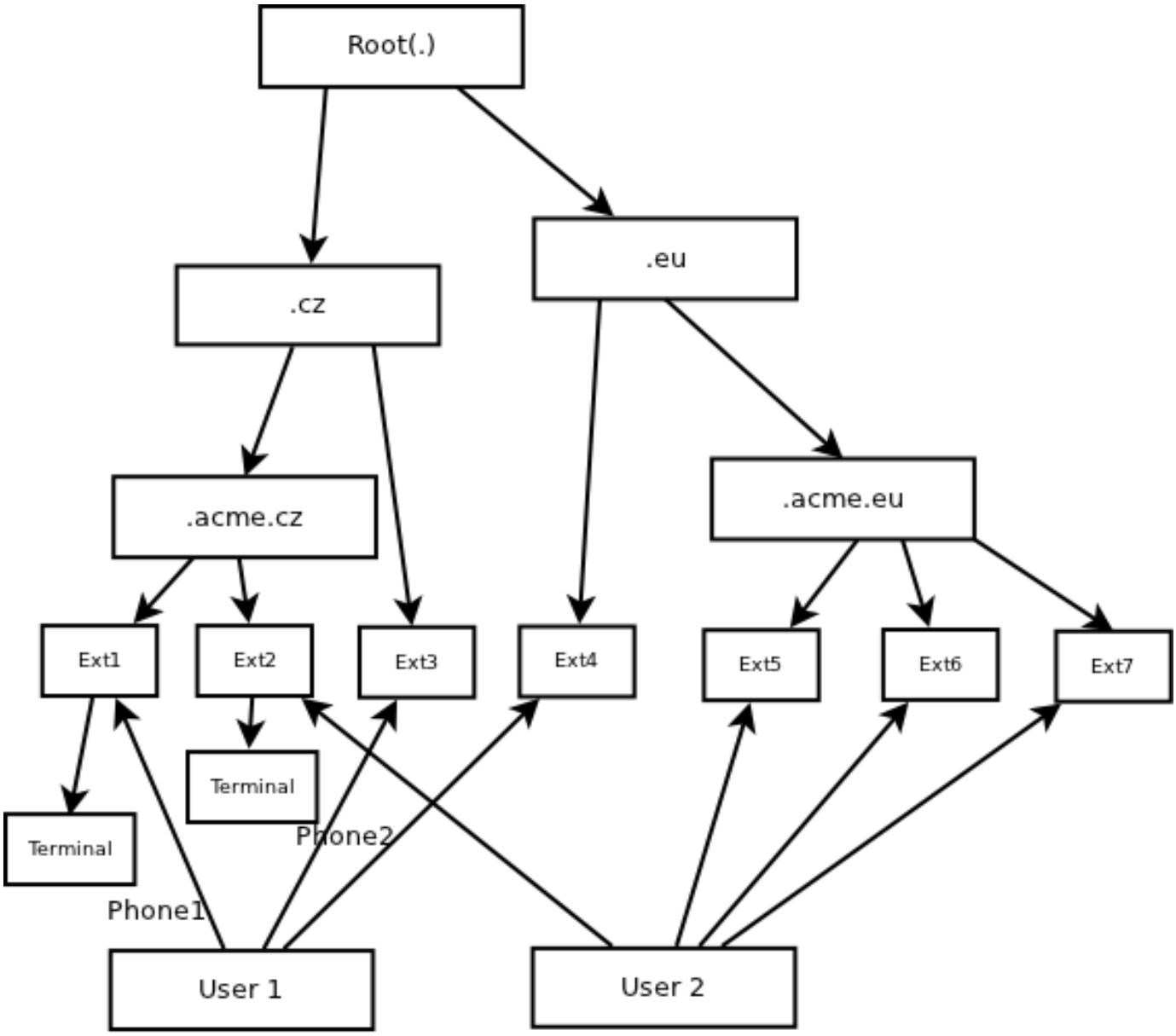


Fig. 1. Container hierarchy

it. According to SIP, secondary engine are defined by DNS SRV and NAPTR records with lower priorities. Engines are identified by unique id which can be DNS name or IP.

B. Voice-Protocol

VoIP protocol is used for call . One engine can serve more protocols. Each protocol has to be described and configured for engine. Complex protocol settings are out of scope of this article. But should be placed into engine protocol context. Each engine can have more voice protocols, for example Asterisk. Protocols used in this document are *SIP*, *H.323*, *MGCP*, *IAX*. But it can be easily extended to more protocols.

C. Signaling-Transport

Signaling used by Voice-Protocol. Some protocols can divert signaling and media (like SIP) and other cannot (ISDN). In this container, we define transport parameters for signaling. There can be more enabled signaling transports per protocol if it is supported. For SIP, this container can contain *UDP*, *TCP*, *TLS*. For ISDN, it is always *D* channel.

D. Media-Transport

Media transport mechanism for voice protocol. Each voice protocol can use some kind of media transport. For SIP, this can be RTP or SRTP. For ISDN, it is *B* channel. Here we can define even allowed codecs and codec preference for each protocol.

E. Root

Root is upper level container. All containers options are inherited from from it. It has special meaning because it is gateway to outside world. Even if it is possible to link any external system to any part of tree, this is main container, which is used for all system, not specified here. See listing 6. Root container must have at least one primary engine and protocol.

Listing 6. Container

```
container . {
 engine kamailio at 10.1.1.1 as primary {
  voice-protocol sip { ... }
 }
}
container .cz {
 engine asterisk at a.b.c as primary {
  voice-protocol sip { ... }
 }
}
```

F. Non-root Container

Containers are glued together in hierarchical graph. Each non-root container must have one superior container and zero or more inferior containers. Each container must have at least one primary engine and protocol. There can be any number of users, extesions and options in each container. Each container can have its own protocol subcontainer, which defines protocol options common to all inferior containers.

G. User

User which use system. Can be employee, customer or administrator. One user can have more extensions associated. User can exist in any container. User defined in superior containers are automatically inherited to inferior containers. For example, user defined in .cz container will work for any subdomain. This eases configuration for huge systems, where users with higher priorities can manage or use subsystems automaticaly. Each user can have static or dynamic variables associated with him (like email or call credit). See 7.

Listing 7. User options

```
user boss {
  secret abcdef
  name John
  %email a@b.c
}
```

H. Extension

Each user can have zero or more extensions. Extensions can have any ASCII represenatation, however, some engines do not support non-numeric extensions. Extensions inherit options from user context. For example, if user has email globaly defined, this is propagated to all his extensions options unless explicitly specified for extension. Each extension can have static or dynamic variables associated. See 8.

Listing 8. Extension options

```
user boss {
 extension 123456 {
  protocol SIP
  authid user_id
  secret passwd
 }
}
```

V. Call routing

Primary function of VoIP system is of course call routing. Call routing can be very complex. In optimal situation, we do not need any extension mangling system. Instead of this, we should call anybody with entire uri (user part and domain part) and routing will be made based on DNS SRV and NAPTR records. But world is more complex that it seems to be. Most people in world uses classic telephones and telephone numbers. It will not change in near future. For example, routing scenario for classic telephones is another than classic SIP routing. When using SIP, we take domain part into account, but in telephone world, we take care only about called number (user part). We have to define this mapping somewhere and this is main goal of this section. Next function of call routing is to allow/deny calls or change call parameters based on source container, destination container, source extension and destination extension. At this point, it is good idea to be inspired by IP world. Comparing to it, we can use similar mechanisms to route voice calls. Inspiration for this syntax is BSD packet filter because it is clear to read and very easy.

A. Call as packet

When comparing to IP world, we can look into call as IP packet. There are some informations inside which are used to make decision, what to do with packet. It can be dropped, passed, translated, redirected etc. It is very similar to call. In our situation, source IP address is caller and destination IP address is callee. So we can create rules, which will be evaluated one by one until matched and after it, somehow changed. Instead of IP, we have much more informations about call, so rule can look more complex. But when configured, we have good routing factory which will be adapting to each call, not just context or protocol.

B. Basic syntax

Basic syntax of call router is similar to BSD packet filter. Instead of IP and subnets, extensions and wildcards are used. Config uses selectors, wildcards, actions and action options. Selectors will look into call parameters and if all of them on line are matched, action with wanted options is fired. See example of main routing syntax in 9.

C. Actions

There can be variety of actions but let us explain only some of them:

route Route call without any change of its media. Only caller and callee information can be changed. This is like SIP proxy scenario

proxy Proxy call from one side to another. Like media gateway

redirect Redirect call

callback Make a callback
ivr Jump to IVR menu

D. Action options

Each action can have many fine-tuning options. They are defined in action options block. There can be many tunng options like maximum number of calls, changing call codecs, changing src or dst number etc.

E. Wildcards

*	Any extension with any context
@ctx	Any extension in context ctx
12*ctx	Extension starting with 1 or 2 and any number of characters after it
....@ctx	Any extension with four characters
(.*)@ctx	Regular expression. Not all systems support it

F. Selectors

protocol	SIP, H.323, ISDN, IAX
engine	Engine which entered or destined call
from	#extension, @user, wildcard or regular expression
to	#extension, @user, wildcard or regular expression

Listing 9. Basic syntax of call router

```
container acme.cz {
 action selectors [wildcards] {
  [action options]
 }
}
```

VI. Practical scenario

In practical world, entire config system should work as parser between user and system. User enters commands and parser will change configuration of routing engine according entered parameters. There are three levels of parsing which should be implemented sequentialy. First, there will be YANG schema definition and simple config parser, which takes YANG input configuration of system and produces output for routing engine. This is one-way process from high level config file into group of files suitabled for selected routing engine. Next, CLI parser will be created. It will be more complex because it should interact with user with autocompletion and interactively change parameters of entered context (needs better state-machine). Third level is to try to make reversed scenario, so parser or CLI process should ask system for parameters in interactive mode. This part is most complex because it needs ability to get states and config data from target system interactively.

A. Asterisk

Asterisk as routing engine will be configured creating text config files under /etc/asterisk directory. In first and second stage, it is enaugh to create only config files up to second stage. But in third stage, database with states and some kind of RPC into Asterisk will be needed.

B. Kamailio

Kamailio parser will produce only one config file with entire routing logic. For small systems, it can do entire routing based on Kamailio configuration script. But for more complex systems and for third stage, database and RPC is needed to communicate both ways.

C. Cisco

Cisco parser will produce only one config file with startup-config. Exported text can be only part of Cisco config, which is responsible for VoIP routing. Text file should be created according to Cisco Unified Border Element configuration syntax.

VII. Conclusions

This script tries to define new syntax. It is not so easy to do so. There is still lot of work to tune all parameters and options to be strict and clear. But this is very good point of start. Now we are working on another project, where we want to use this script functions in practical situation. During this stage, there will be many improvements which will be finded just in time when it occurs. When definition of syntax enters stable version, we will try to write RFC and document everything. We want to use as much existing RFCs and libraries as possible.

References

[1] MEGELEN, Jim, SMITH, Jared, MADSEN, Leif. Asterisk: The Future of Telephony. Sebastopol: O'Reilly, 2005. ISBN: 0-596-00962-3.

[2] Voznak,M. SIP as a Framework for Developing Communication Systems. Vyzvana prednaska na Fakulte elektrotechniky a informatiky Slovenske Technicke Univerzity v Bratislave, Bratislava 8.12.2010

[3] Voznak,M., Rezac, F. Rozhon, J. "Applied multiprotocol routing in IP telephony, "Journal AEEE - Special issue 2010, Volume 8, Number 5, pp. 118-123, December 2010, ISSN: 1804-3119

[4] Network configuration protocol (RFC 4741)

[5] YANG Module for NETCONF Monitoring (RFC 6022)

[6] YANG - A Data Modeling Language for the Network Configuration Protocol (RFC 6020)

Implementing WebSockets in Web applications

Jiri Vychodil, Filip Rezac, Karel Tomala, Miroslav Voznak
Department of Telecommunications
VSB – Technical University of Ostrava
Ostrava – Poruba, Czech Republic
jiri.vychodil@vsb.cz, filip.rezac@vsb.cz, jan.rozhon@vsb.cz, miroslav.voznak@vsb.cz

Abstract— Customer management system (CRM) is information system which is very widespread in corporate sector. The good relationship with customers is very important in this area. For its improvement a connection CRM with phone branch exchange is used: if a client of phone exchange is calling employee whose number is set in CRM, CRM registers the call and immediately informs employee about all relations with the client. Present internet technologies allowing this feature use connection of JavaScript and classic HTTP communication. Indeed preferable from all sides is to use WebSockets – technology for bidirectional, full-duplex connection intended for use between web browser and server. Our work deals with comparison of technical intensity of both technologies and practical use of WebSockets.

Keywords-component; VoIP, WebSockets, CRM, HTTP, AJAX, JavaScript, Asterisk

I. INTRODUCTION

Web application is software offered to users of internet from web server through computer network – inter or intra net. These applications are so popular, because web browser as a client je accessible almost everywhere. Application is run on the server and web browser is used only as "thin client", because the browser itself does not know the logic of application. As the most known example of web application can be considered web interface of mailbox.

Today in the field of web application for communication between a client and a server is used AJAX (Asynchronous JavaScript and XML) [1]. So the asynchronous client-server model of transfer is used. If a user runs an application (for example set "create a new email" in mailbox), JavaScript set XMLHttpRequest query to the server and it responses through the same protocol and form of new email appears.

But if thin client needs to find out if a new event occurs (new income email) which the user should know about, the thin client has to periodically query the server. A TCP connection must be created again (Transmission Control Protocol) for each query separately, which increases the requirements for capacity of network transmission, network elements and server, where the application is running. So in our example the thin client periodically checks if there is new incoming email in mailbox. Let us assume, that for the comfort of the user that every one minute the TCP connection and query are set. If there is no new email – as usually – this communication is needles. Indeed in the case of announcement of new incoming call one minute is too long period and the checking must be done in periods at most several seconds long.

Due to disadvantages of this classic communication which is used due to historical reasons, because there is no better way, W3C is developing WebSockets [2] – technology for asynchronous bidirectional full-duplex communication over a single Transmission Control Protocol socket.

II. USED TECHNOLOGIES AND SOFTWARE

We had aimed at use of WebSockets in our work. As mentioned above, this connection is based on TCP socket between web browser and a server. Both browser and server have to support the technology.

JavaScript [3] is responsible for building the connection as well for the communication through WebSocket on the browser side. The client sends handshake request to start the connection and the server sends a handshake response.

WebSocket handshake request:

```
GET /demo HTTP/1.1
Upgrade: WebSocket
Connection: Upgrade
Host: example.com
Origin: http://example.com
Sec-WebSocket-Key1: 4 @1  46546xW%0l 1 5
Sec-WebSocket-Key2: 12998 5 Y3 1  .P00
^n:ds[4U
```

Server handshake response:

```
HTTP/1.1 101 WebSocket Protocol Handshake
Upgrade: WebSocket
Connection: Upgrade
Sec-WebSocket-Origin: http://example.com
Sec-WebSocket-Location:
ws://example.com/demo
Sec-WebSocket-Protocol: sample
8jKS'y:G*Co,Wxa-
```

The technology of WebSockets is still in the phase of developing and W3C has not published the final draft. Only part of the complex problematic of WebSockets has been introduced –

draft-ietf-hybi-thewebsocketprotocol-00. This document is outdated, however it suffice for the purpose of a demonstration.

The handshake looks like HTTP but actually it is not. It allows the server to interpret part of the handshake request as HTTP and then switch to WebSocket. Sec-WebSocket-Key1a 2 the server uses to construct a 16-byte token at the end of its handshake to prove that it has read the client's handshake.

We have used web browsers Chrome 10.0.648.204 and Opera 10.01 in our work. Because some mistakes were found in the proposal of WebSockets, which could be security threat, WebSockets are not permitted by default and must be allowed. Another browser supporting sockets is Safari, but because new versions of modern browsers are released very often, testing of all browsers was not out priority.

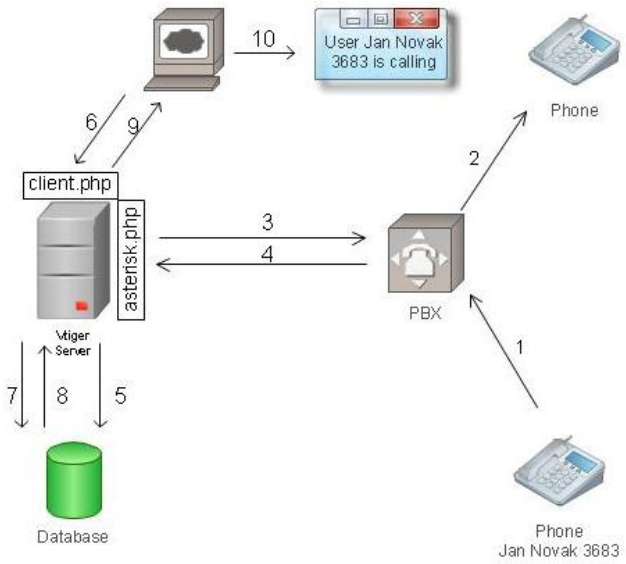


Figure 1. Solution without using websockets

Vtiger [4] is a free Open Source CRM (Customer relationship management) software. It is written in language PHP and use MySQL database [5,6]. As other CRM systems it provides many possibilities for relationship, but we have aimed at features, which are closed to telephony. Vtiger allows its users to assign a phone number to all its human entities. A connection of Vtiger and PBX Asterisk [7] through management account in Asterisk brings much more possibilities and features. Vtiger is thus connected with PBX through TCP connection a monitors all events like incoming call to user of the system and notifies him of incoming call in his web browser. The user can see not only who is calling but he can also display all information about the caller like trade made, transactions, orders, amount waiting to be paid and all this by one click

III. PRESENT SOLUTION

The present solution of notifying a user of incoming call. The user is logged in a web browser into the CRM Vtiger. Caller makes a call through the Asterisk (1), which redirects the call to the called (2) and its phone starts to ring. PHP script asterisk.php of Vtiger monitors the Asterisk and founds out, that the number of inner user is called (3, 4). The script saves this information into MySQL database (5). JavaScript running on the side of the inner user sends query to script client.php (6) periodically to ask, if there is any incoming call saved in the database (7,8) and sends the answer (9) to the browser. In case the information of incoming call exists, the user is informed about the situation with pop-up window in his web browser (10).

This solution uses a very high amount of TCP connections and database queries. After the use of packet analyzer we found out, that 2 kilobytes were transmitted during one TCP connection created for one query from client to the server. Taking to the consideration, that the number of connected clients is 40, we can count, that 4800 TCP connections are created and closed during one hour, as well 4800 inquiries are submitted and 10 MB of data are transmitted.

This has to be because JavaScript has to ask a server with the use of HTTP periodically, there is no way how a server could send any information to the browser without browser's asking.

IV. SOLUTION WITH USE OF WEBSOCKETS

Number of created TCP connection between client and server is largely reduced with use of WebSockets as well as volume of data. This is possible because WebSocket connections is full-duplex and both directional, so server does not have to wait for client's query, client does not have to send queries periodically, but the server sends any information without the query. In addition the time between the phone starts ringing and user is informed who is calling is decreased. This time was up to 3 seconds in the worst case and now we can say the time is insignificant.

After the user logs in to Vtiger system a JavaScript is started. This script creates WebSocket connection with script asterisk_ws.php run on the server. That script services all sockets (listen on set TCP port, creates and keeps sockets). It also has to know which users are logged into the system and theirs phone numbers. A use of database is not suitable, as the script would query the database, which would increase the computational demand. Thus JavaScript sends the user's phone number after the logging and script asterisk_ws.php knows them, although it does not save them anywhere.

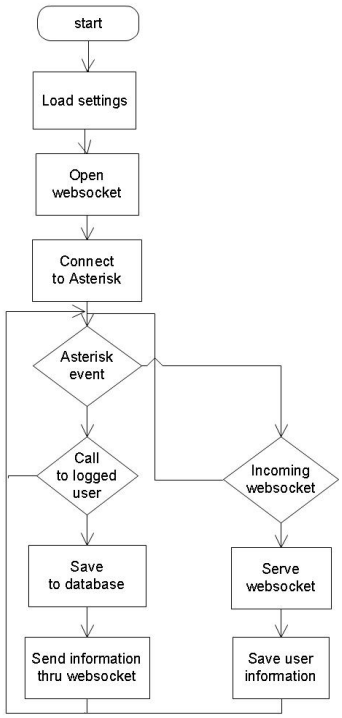


Figure 2. Schema of function asterisk_ws.php script

A schema of rebuild system is shown on the figure 3. After the creation call to the Asterisk (1) the call is redirected to the called party (2). Script asterisk_ws.php notices this event from extract of Asterisk management account (3, 4) and if the called number is one of system users, it saves this information to the database (5) for archiving purposes. If the called user is logged the script sends him information about caller – phone number and a name (7). Pop-up window with this information then appears in called user's web browser and the user can show caller's information by one click.

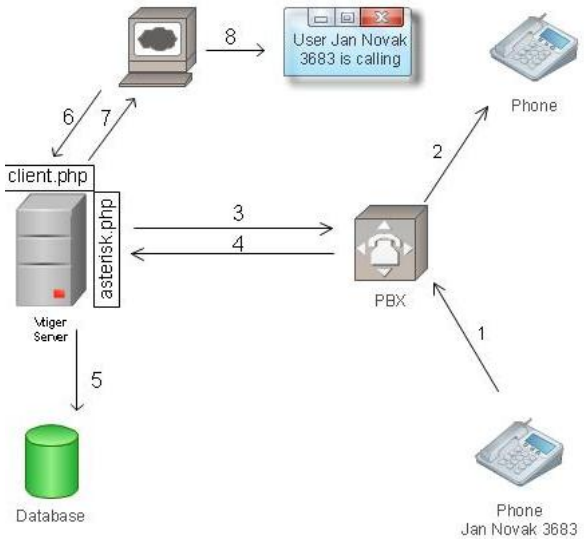


Figure 3. Scheme of rebuild system

A. *WebSockets advantages*

Advantages of WebSocket are then obvious. Moreover if the whole information system is written with use of AJAX creating only one socket is necessary - after logging to the system. If not, the socket is created again with every changing of shown page. Rewriting all system so that all communication would use sockets is possible. Then the transmission would use the socket not only for sending information about calls but also for transfer content of pages. The amount of transferred data would not be decreased, but the number of needed connections would.

B. *Settings of Asterisk*

To create management account in asterisk is necessary modify file /etc/asterisk.manager.conf by adding these lines:

```
[general]
enabled = yes
port = 5038
bindaddr = 127.0.0.1
[vtiger]
secret=vtiger
deny=0.0.0.0/0.0.0.0
permit=158.196.244.0/255.255.255.0
read=system,call,log,verbose,command,agent
,user, originate
write=system,call,log,verbose,command,agen
t,user, originate
```

We have allowed management account by this and set the Asterisk to listen on TCP port 5038 for incoming connection. The name of account is "Vtiger" with the same password set in item secret. Range of allowed IP addresses which are allowed to create incoming connection is set the laboratory one for increasing the safety level. If users of Vtiger use Asterisk, it is necessary to create account for them. In the other case it is not.

Next usage of WebSockets in Vtiger system is possible for creating a call from web interface, when user clicks the name of called user and the JavaScript send the command to the system and it resends the command to Asterisk. It first creates a call to caller and after he answers the call, Asterisk retransmits the call to called. WebSocket is used for transmitting this order, there is no need to show an schema. The setting of Asterisk has to allow to management account creating calls with word "originate" in items read and write. We have to notice, that this is not mentioned in online Vtiger install manual.

C. *Setting of Vtiger*

Setting of Vtiger is done through web interface. It is necessary to assign phone numbers to users and set up a connection with Asterisk. Information about IP address of Asterisk, TCP port, name of management account and password and version of Asterisk (1.4 or 1.6) is needed. Next error in Vtiger appears here, when it is not possible to set up version of Asterisk and logging in to database of Vtiger and setting it by MySQL query is necessary. After this, the web interface runs properly and we can choose between versions.

V. FUTURE WORK

WebSockets are still in the phase of development and W3C still is still publishing new drafts. It is necessary to wait for final version to use sockets in real service. Because there was a mistake in drafts in the previous versions, it will be necessary to change handshake messages on both sides – the client and server one. Implementation of new draft in web browsers can also take some time.

Better support of network elements can be another improvement. There is no problem in a case of using WebSockets in an area of one network domain, but if a user use proxy for accessing the internet, the connection probably fails. One possible solution of this situation is to use Encrypted WebSocket connections called WebSockets Secure. This connection uses TLS (Transport Layer Security) and WebSockets is created above this connection, so proxy does not need to support sockets.

VI. CONCLUSION

WebSockets are relatively new technology, which je still in the phase of development, but useable applications can be created, although at the expense of future rebuilding caused by the fact, that the final draft has not been published.

Sockets are quite a new technology and easy way of both directional full-duplex permanent connection between a client – web browser and server. They offer a communication, which other technologies do not, nowadays the most used solution is AJAX, which has to create a TCP connection and set HTTP query every time a web browser transmits a data.

Our work deals with implementing WebSockets and integrating them into open source CRM system Vtiger. After this upgrade the requirements to client, server and network elements and the total amount of transmitted data rapidly decreased. Also time needed to inform the user about the person who is calling is now insignificant.

ACKNOWLEDGMENT

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 218086 and FRVS G1 1672 /2011.

REFERENCES

[1] B. Brinzarea, . Darie, F. Chereches-Tosa, M. Bucica, “AJAX and PHP: Building Responsive Web Applications,” Packt Publishing; 1st ed., March 2006. ISBN-13: 978-1904811824

[2] The WebSocket API, URL: http://www.w3.org/TR/websockets/

[3] Vtiger CRM, URL: http://vtiger.com/

[4] D. Flanagan, Javascript: The Definitive Guide, Third Edition, 3rd ed., O'Reilly Media June 1998

[5] NIXON, R. *Learning PHP, MySQL, and JavaScript,* 1st ed. O’Reilly Media, 2009.

[6] DUBOIS, P. *MySQL Cookbook*. 2nd ed. O’Reilly Media, 2007. s. ISBN 978-0-596-00145-2.

[7] J. Meggelen, J. Smith and L. Madsen, "Asterisk The Future of Telephony," O'Reilly Media; 2nd edition, 2007

Design of Digital Communication System with use of RF VSG PXI-5670 and RF VSA NI PXI-5661 in role of Software Defined Radio

Radek Martinek, Jan Zidek
Department of Measurement and Control
VŠB – Technical University of Ostrava
Ostrava, Czech Republic
{radek.martinek.st1, jan.zidek}@vsb.cz

Karel Tomala
Department of Telecommunications
VŠB – Technical University of Ostrava
Ostrava, Czech Republic
karel.tomala@vsb.cz

Abstract - **In the recent days, the needs to transfer more data by more users is increasing dramatically. In association with this fact comes to the foreground the efficiency of bandwidth utilization, as the bandwidth is the determining limit factor for growth of this requirement. In the last years in order to ease the pressure of this necessity, the modulations were transformed from simple analog (AM, FM and AM) to simple digital modulations (ASK, FSK, PSK) and complex digital multistate modulations (M-PSK and M-QAM).**

The authors are focused on evolution of particular stages in the development of modern communication systems as seen from the perspective of digital modulations (4PSK, 8PSK, 16PSK 32PSK, 4QAM, 8QAM, 16QAM, 32QAM and 64QAM). These digital modulations are investigated with the eyes on the design of digital communication system (efficiency in utilization of bandwidth, of power and cost). The probability of wrong input for the selected modulations was theoretically calculated. These results were compared with the direct measurement of bit error rate (BER) while using the PXI modular systems and LabVIEW development environment.

On real measurements using the vector signal generator RF VSG NI PXI-5670 and vector signal analyzer RF VSA NI PXI-5661, the authors aim to acquaint the readers with not very widespread synthetic devices in the application area of wireless transmission systems. The paper describes the generation and analysis of digitally modulated signals using hardware platform as PXI - VSG, VSA, and library functions from the additional library of Modulation Toolkit.

Keywords - RF VSG NI PXI-5670; RF VSA NI PXI-5661; BER; Synthetic (Virtual) Instrument; Communication Systems; Flexible Measuring Systems, Modulation Techniques, Modulation Toolkit.

I. INTRODUCTION

When designing a digital communication system, the focus of the design is usually in one of those three areas:

- *The efficiency of bandwidth utilization* describes the ability of the communication system to transfer the required data in limited bandwidth.
- *The efficiency of power utilization* describes the ability of the communication system to transfer the required data with as small power consumption as possible.
- *The costs efficiency* describes the ability of the communication system to transfer the required data with the lowest costs incurred to build the system.

Let's look to the different weighing of the three factors in the design of various communication systems:

1. The design of the terrestrial stationary microwave communication system will focus on the efficient use of frequency bandwidth and low error rate of the system. Since it is a stationary system, there is no problem to supply a plenty of power. Furthermore, once there will be just a limited number of these communication systems, the acquisition cost of the technology does not play the main role either.

2. While designing the mobile phones for cellular networks, the power efficiency becomes the major problem, while the limited battery capacity of mobile device defines the running time. Also the costs efficiency plays the important role as the customers look for the best price when buying a mobile phone.

The relation between these three factors is given by the fact that improving of one of them usually leads to making the other factors worse. For example, the attempt to the most efficient bandwidth use leads to complex treatment of hardware with higher price. Minimizing the cost of product is a general requirement, which helps to promote the technical solutions on the market. However, if the price reduction was reached with a technical solution that leads to increased power consumption or reduced efficiency of bandwidth utilization, the final product would have also reduced chance to compete on the market.

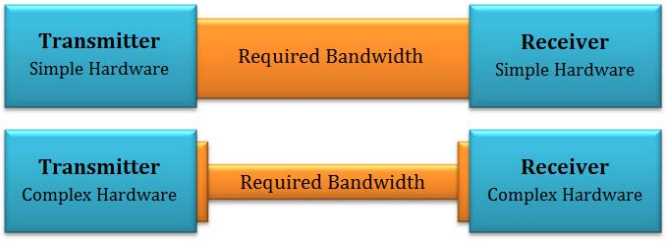


Figure 1. Dependence of required bandwidth on the complexity of hardware

The bandwidth needed for transferring the data volume is inversely proportional to the complexity of the used hardware. Simple hardware on the side of transmitter and the receiver usually leads to the need of higher bandwidth for the system running, while the complex hardware of the transmitter and receiver allows the use of sophisticated modulation schemes and thus the transfer of the data while using a smaller bandwidth (Fig. 1).

This article presents the general concepts of a flexible measurement system designed for generation and analysis of digitally modulated signals.

II. Probability of Error Receive

Frequency bands are limited and the number of users with the need to use them is constantly increasing. Multistate digital modulation allows transferring much more information in the bandwidth than simple digital modulation.

In theoretical calculations the additive white Gaussian noise AWGN [3] is assumed as the only source of interference. However, the probability of error receive is actually greater. This may be caused by the following factors. First, the imperfect frequency filtering in the transmitter or receiver. Second, the phase noise of oscillators, which are used for the generation of modulated signals and non-linear amplifier. Third, the wave propagation for multiple paths and the associated types of leaks. The real error rate could be approximately reached by considering all of these interferences while calculating the error probability, but in most cases the error rate is detected by direct measurement of bit error rate (BER) [12]. Noise is considered as a random signal, while it is not possible to determine its value at the random moment.

The formulas for the theoretical calculating of error rate probability are usually expressed by complementary error function erfc(x), complementary distribution function Q(x) or distribution function F(x) [6]. The conversion between the most widely used erfc(x) and Q(x) function is expressed by the following formula [7]:

$$Q(x) = \frac{1}{2} erfc\left(\frac{x}{\sqrt{2}}\right) \tag{1}$$

In formulas there is a variable ε (2), which represents the average transmitted energy spent per bit, relative to the average energy of noise [1]. This value corresponds with standardized ratio E_b/N_0 (energy per bit to noise power spectral density). The dimensionless quantity E_b/N_0 is independent on the modulation method and expresses the rate ratio of the useful signal in the received signal [6], [5]:

$$\varepsilon = \frac{E_b}{N_O} \tag{2}$$

E_b - E_b – mean value of energy per bit, N_o - power spectral density of noise.

A. PSK Modulation

The resulting formula of error probability for PSK is the same as for QPSK modulation [6]:

$$P_e = \frac{1}{2} erfc\left(\sqrt{\frac{P^2 T_b}{2N_0}}\right) = \frac{1}{2} erfc(\sqrt{\varepsilon}) = \\ = \frac{1}{2} erfc\left(\sqrt{\frac{E_b}{N_0}}\right) \tag{3}$$

P_e – probability of bit error, P - mean power of signal, T_b - time duration of binary symbol. For multistate variants of keying the following formula for the error rate of symbols is valid then [2], [4], [6]:

$$P_s = erfc\left(\sqrt{\frac{E_s}{N_0}}.\sin\left(\frac{\pi}{M}\right)\right) \tag{4}$$

Ps – probability of symbol error, Es - mean value of energy per symbol, Es / No - standardized signal-to-noise ratio related to the energy per symbol. For the bit error rate the formula (5) is actual. more detail in [6]:

$$P_e = \frac{1}{\log_2 M} erfc\left(\sqrt{\frac{E_b \log_2 M}{N_0}}.\sin\left(\frac{\pi}{M}\right)\right) \tag{5}$$

B. QAM Modulation

While using the M-QAM modulation, during the transmission by channel with AWGN noise or during coherent demodulation, the probability of error in the transmission of symbols is determined approximately by formula (6), see [2] [3], [6].

$$P_e \cong 2\left(1 - \frac{1}{\sqrt{M}}\right) erfc\left(\sqrt{\frac{E_0}{N_0}}\right) \tag{6}$$

III. Synthetic Measuring Instruments

The philosophy of synthetic (virtual) measuring instruments is very progressive, since it allows to create the instruments, whose software-implemented functions exactly match the requirements of users, while keeping the performance parameters of classical measuring instruments. The functions are defined by the user. The role of key component is assumed by the software in this case [12]. Figure 2 shows a difference between the traditional and the synthetic measuring instrument. Red color indicates the key components that make up the foundation of the measuring instrument.

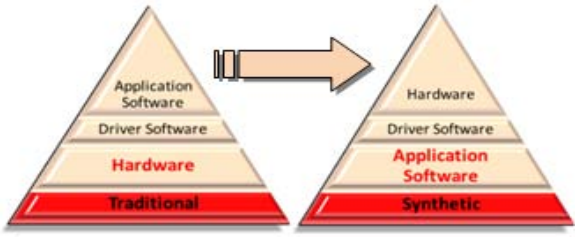


Figure 2. Different concepts of traditional and synthetic (virtual) measuring devices

Architecture of traditional and synthetic instruments has similar hardware components. The main difference between these types of architecture is in the location of the software and its available to users. Traditional or separate instruments have all the components placed in one box for each device individually. The measurement functionality, analysis, displays and control devices are specified by the manufacturer. The user is therefore limited to the special purpose of measurement device and it is very difficult to extend the capabilities of it. In contrast with such device, a software-defined synthetic instrument provides access to the raw data from hardware to users who can define their own measuring and user interface, more detail in [1], [12].

IV. Concept of Measuring Equipment in the Form of Synthetic Instrumentation

In recent years, the number of new protocols used in digital transmission systems is increasing dramatically. Some of them are completely new, others are gradually replacing existing ones. The introduction of digital transmission systems of the new generations, accompanied with new and new standards, evokes the need for flexible test platform, both in their development stage and in their use. The authors of this article presented a solution in the form of a test system based on the synthetic instrumentation. The concept of the measuring apparatus is shown in Figure 3.

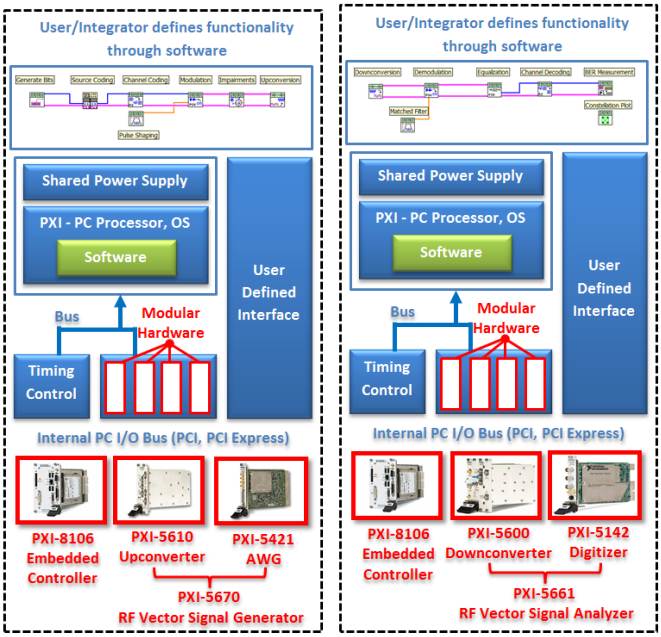


Figure 3. The concept of measuring equipment in the form of synthetic instrumentation: RF Vector Signal Generator and PXI-5670 RF Vector Signal Analyzer PXI 5661

One of the biggest advantages of this test platform is the ability to respond flexibly to new requirements in the area of testing. In this article the RF vector signal generator and PXI-5670 RF vector signal analyzer PXI-5661 have been used. These two modules are highly flexible solutions for testing of new digital transmission systems. The roles of software and hardware are shown in Figure 4.

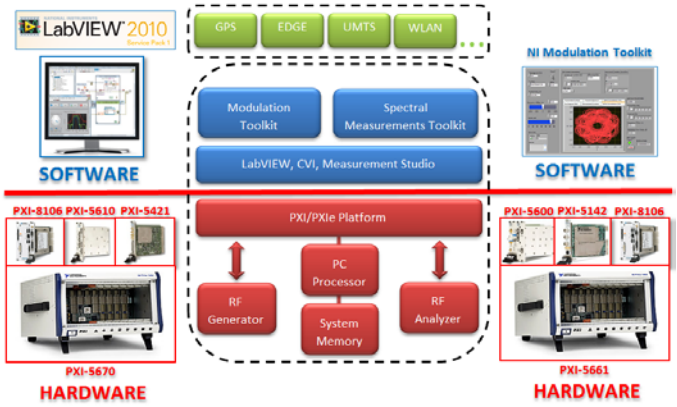


Figure 4. The role of software and hardware in the utilized test system

A. *Vector Signal Generator - NI PXI 5670*

Along with the development environment LabVIEW, complemented by extended Modulation Toolkit library, the PXI module could be used to generate the required test signals for verifying the possibilities of digital transmission systems which use the new standards. The whole test system can be easily adapted to new requirements, while the focal point of functionality of such a system is located in the software part (Figure 4).

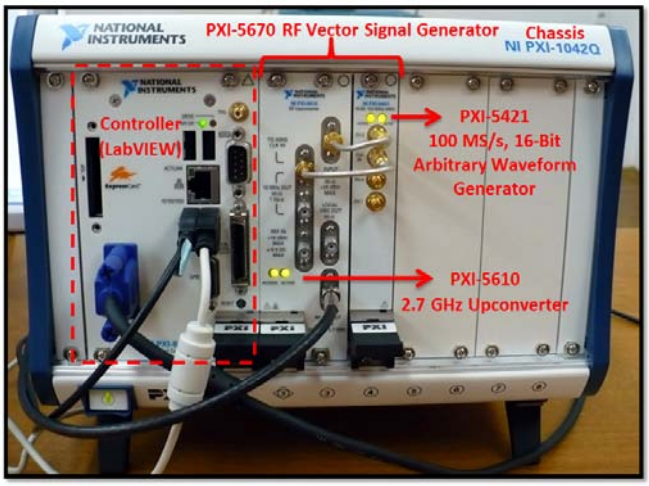


Figure 5. Real scheme of NI PXI 5670 - Vector Signal Generator (NI PXI-5610 – RF Upconverter, NI PXI-5421 -100MS/s AWG)

RF vector signal generator, the NI PXI-5670, represents the generator of user-defined waveform (arbitrary waveform generator) working a resolution of 16 bits and sampling rate of 100MS / s (400MS/sv in the interleaved mode) with a depth of memory up to 512 MB and the real bandwidth of 20 MHz. Using a digital upconvertor along with this module can generate signal in the range of 250 kHz to 2.7 GHz with random modulation scheme such as: AM, FM, PM, ASK, FSK, MSK, GMSK, PSK, QPSK, PAM, and QAM, see [10].

B. *Vector Signal Generator*

For the analysis of digitally modulated signals a NI PXI-5660 module was used. This module is a very compact solution

(30% of normal weight and cubature of separate devices in this class), allowing very rapid measurement of digitally modulated signals in the range from 9 kHz to 2.7 GHz. With the real bandwidth of 20 MHz, but with possible flow of data 132 Mb/s over the PCI bus, this solution represents tremendous progress in the contrast of 1 MB throughput via GPIB interface for connection of single vector signal analyzers, see [11].

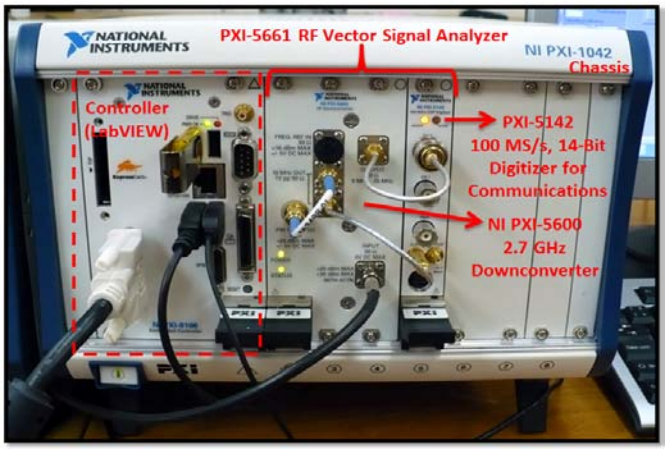


Figure 6. Real scheme of NI PXI 5661 - Vector Signal Analyzer (NI PXI-5600 – RF Downconvertor, NI PXI-5142 -100MS/s OSP Digitizer)

Along with the development environment of LabVIEW with Modulation Toolkit extension libraries and Spectral Measurement Toolkit, this module represents very flexible platform for the automation of usual parameters measurement such as: in-band power, adjacent channel power, power and frequency-peak-search. Visualization of measurement results is possible in traditional forms such as 3D spectrograms and constellation diagrams for analysis of digitally modulated signals (I/Q Modulation for Data Analysis).

V. Verification of Designed Transmission Wireless System Functionality

The data, which are transmitted from one user to another, must pass through a transmission chain composed of several important basic blocks (Figure 7).

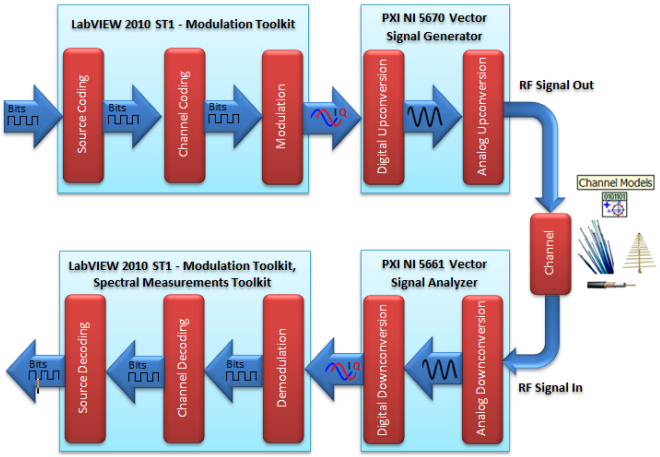


Figure 7. Block diagram of the realized digital transmission chain

In the block of channels three aspects have been used: a simulated channel, interconnection by the cable and real wireless connection. The verification of functionality of designed system was done in two ways. First, the ideal signal was generated, in the second part of measurement, the degradation of signal in the form of MER was added to the generator. The signal to noise ratio was influenced by MER values with error rate E_b/N_0. In order to eliminate the environmental effects on degradation and transfer of the signal, a 50 Ω coaxial cable was used (instead of antenna) to connect equipment at the first step of authentication.

A. *Evaluation of the Measuring*

Measurements were carried out for all the modulation under the same conditions (i.e. with the same performance would have to be adapted to the spacing between symbols in different modulation), so they would credibly be compared.

B. *Verification of real function of the proposed system*

For verification of the real features of the designed system NI PXI, a passive omnidirectional WiFi Antenna TP-LINK TL ANT2405C with 5 dBi gain – was connected to output of broadcast part of the system and the receiver was affixed with the passive panel sector WiFi Antenna Centurion Wireless Technologies with 9 dBi of gain. In the transmitter the signal was generated by the studied modulations on the carrier frequency 1.96 GHz, power level of 10 dBm signal with a bandwidth 3.84 MHz, Roll factor α of the root raised cosine filter 0.33 and symbol rate 2.625 MS/s. On the receiver side, the quality of the received signal was evaluated by the BER measurement.

Figure 8. Used Antennas (TP-LINK TL ANT2405C and Centurion Wireless Technologies)

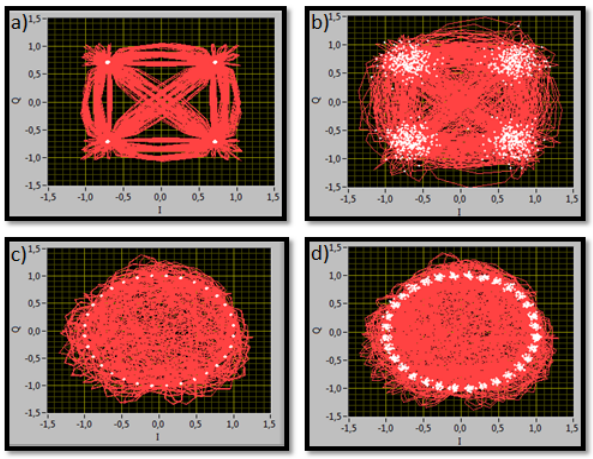


Figure 9. Constellation diagram: a) 4PSK (E_b/N_0=40dB), b) 4PSK (E_b/N_0=10dB), c) 32PSK (E_b/N_0=40dB), d) 32PSK (E_b/N_0=20dB)

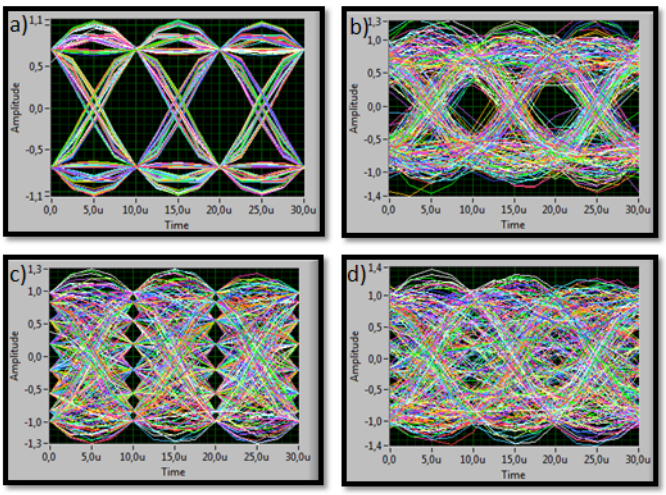


Figure 10. Eye diagram for: a) 4PSK (E_b/N_0=40dB), b) 4PSK (E_b/N_0=10dB), c) 16PSK (E_b/N_0=40dB), d) 32PSK (E_b/N_0=15dB)

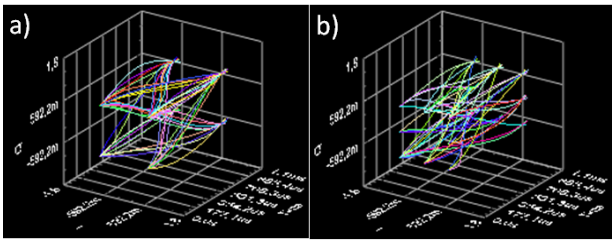
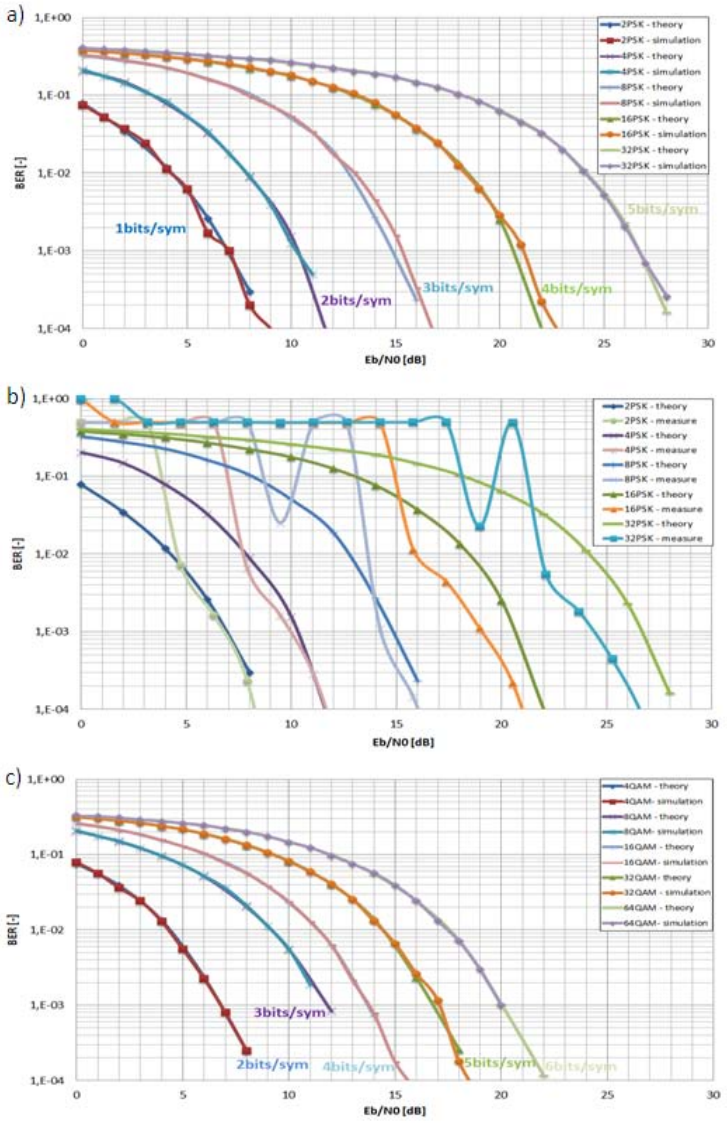


Figure 11. Eye 3D diagram for a) 4QAM (E_b/N_0=40dB), b) 8QAM (E_b/N_0=40dB)

In development environment of LabVIEW individual blocks of transmission chains with AWGN channel for each of the examined modulation techniques were created. From the simulated values the resulting graphs were constructed then. Comparison of calculated and simulated values leads to the same or very similar results, see Figures 12) and c). In Figures 12 b) and 12 d) the real results are compared with the theoretical ones.



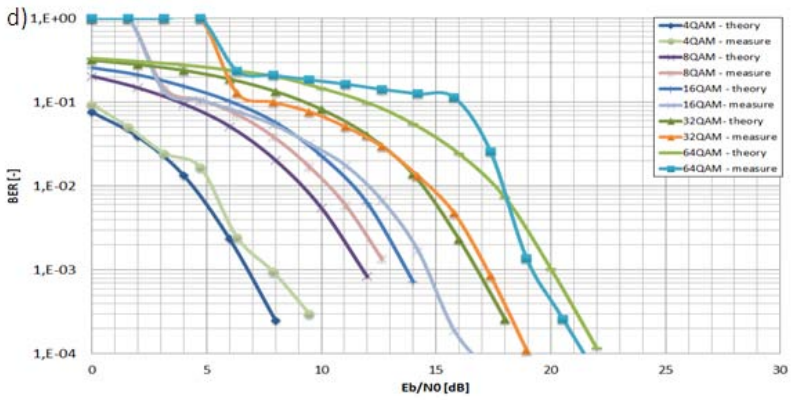


Figure 12. Comparison of graph, depending on the bit error ratio Eb/N0: a) Theoretical and simulated values for M-PSK modulation, b) Theoretical and measured values for M-PSK modulation, c) Theoretical and simulated values for M-QAM modulation, d) c) Theoretical and measured values for M-QAM modulation

VI. CONCLUSION

This article deals with problematic of generation and analysis of signals used for wireless transmission systems. The practical section deals with design and verification of functionality of the system for data transferring using various digital modulations. The main contribution of the work is in the creation of a synthetic instrument designed for real measurement of error rate, applicable in wireless transmission systems. The idea of functionality of such systems comes from the definition of the software radio, which regards the hardware of the transmission system as a universal, generally conceived device. The essential functions of the transfer system are represented by software, which is variable to the users needs. Such a device can be built on the basis of PXI modular system with a LabView development environment, which utilizes flexible programming language G with extensive modular function libraries. On this platform, a functional wireless transferring system was designed and implemented. This measuring system is based on the basic communication chain, see Fig. 7. It contains two separate parts: the transmitter and the receiver. Both of them allow changes in digitally modulated signal parameters, so those instruments can demonstrate the basic principles of wireless transmission for various transmission systems with no necessity of changing the hardware components of the system.

Within the measurements, the theoretical and simulated values of BER for the multistate modulations M-PSK, M-QAM (see Fig. 12 a, b, c, d) were verified. In the terms of design of the digital communication, the flexible nature of modular measuring systems is saving the investments for their acquisition. The software part of those systems, which plays the key-role for the functionality, can be adjusted continuously according to the upcoming technologies in communication and information fields of interest. This aspect will decrease the financial demands on the instrument technologies, which will cover upcoming standards in this area.

ACKNOWLEDGMENT

This post leading to these results has received funding from FRVS F1 498/2011 titled Inovation of Measurement in Telecommunications course and project SP2011/180.

REFERENCES

[1] Martinek, R.; Al-Wohaishi, M.; Zidek, J.: Software based flexible measuring systems for analysis of digitally modulated systems. RoEduNet 2010 9th,Sibiu, Romania 2010, ISSN 2068-1038, ISBN 978-1-4244-7335-9, On page(s):397.

[2] ÖRS, Tolga, CHOTIKAPONG, Yotsapak, SUN, Zhili. Broadband integrated satellite network traffic evaluations. [online]. 2002. Bisante Consortium.
<http://www-bisante.ee.surrey.ac.uk/deliverables/public/bisante-deliverable-22.pdf>

[3] GHASSEMLOOY, Z. Mobile Communication systems – Part 5 Modulation techniques. [online]. 2003. Sheffield Hallam University, Northumbria University.
<http://soe.unn.ac.uk/ocr/teaching/mobile/pp/partV-v1.pdf>

[4] HRDINA, Zdeněk, VEJRAŽKA, František. Digitální radiová komunikace. Praha: ČVUT, 1994, 243 s. ISBN 80-01-01059-7.

[5] HANUS, Stanislav. Bezdrátové a mobilní komunikace. 1. vyd. Brno: T-Mobile CZ a.s., 2005, 134 s. ISBN 80-214-1833-8.

[6] DOBEŠ, Josef. Moderní radiotechnika. 1. vyd. Praha: BEN - technická literatura, 2006, 767 s. ISBN 80-7300-132-2.

[7] PEEBLES, Peyton. Digital communication sytems. New Jersey: Prentice-Hall, 1987, 432 s. ISBN 0-13-211962-5.

[8] PROAKIS, John. Digital communications. 4. vyd. New York: McGraw-Hill, 2001, 1002 s. ISBN 0071181830.

[9] J.S. Wilson, S. Ball, (2009) " Test and measurement" , ISBN:978-1-85617-530-2, USA.

[10] National Instruments: RF Vector Signal Generator NI PXI-5671, katalogový list, 2005
<http://www.ni.com/pdf/products/us/20055170101dlr.pdf>

[11] National Instruments: RF Vector Signal Analyzátor NI PXI-5660, katalogový list, 2005
<http://www.ni.com/pdf/products/us/4mi469-471.pdf>

[12] J. Žídek, (2005) "Grafické programování ve vývojovém prostředí LabVIEW'', Ostrava.

[13] R. Štefan, R. Valrf, (2010) "Solutions in Design, Manufacturing Test'', workshop, Brno.

Robot Navigation System Based on RFID Transponders Integrated with Sensors

Włodzimierz Kalita, Mariusz Skoczylas, Mariusz Węglarski
Dept. of Electronic and Communications Systems
Rzeszów University of Technology
Rzeszów, Poland
wkal@prz.edu.pl

The integration of different kind of physical quantity sensors with transponders of RFID (Radio Frequency IDentification) identifiers (ID, tag) enables to store information about local environment in built-in memory [1]. The quantities (temperature, humidity, pressure, light intensity, vibration, etc.) can be measured with different frequency and in time periods with various duration and the data gathered by the tag can be read at any time by superior control system. A deployment of some number of such equipped identifiers in a particular region gives the opportunity to determine precisely the parameter values describing ambient conditions. This knowledge is very valuable during routing process of moving objects, for example in navigations system of mobile robots especially with high autonomous intelligence. The authors' effort has focused on the subject of sensors integration with transponder IC (Integrated Circuit) circuit and on determining the characteristics for such solution especially in aspect of memory organization. Selected problems connected with autonomous navigation systems of static objects as well those moving inside of defined space have been also presented in the paper.

Keywords - RFID, identification, sensor, autonomous localization, mobile object, navigation

I. GENERAL INFORMATION

The integration of solutions from different fields of electronics, in a single hybrid structure, allows for a significant increase in functional capabilities of electronic systems [2]. The achievement of greater functionality in the single element enables the realization of innovative applications in the area of artificial intelligence creation. In such kind of complex structure, technological characteristics of manufacturing process of individual component as well as integration possibilities are very important issues. The hybridization of a RFID transponder (ID, identifier, tag), sensors of various physical quantities and a data memory of environment parameters on a one shared substrate (usually flexible but also ceramic, PCB, etc.) is the great example of this problem. The general idea of such intelligent identifier with sensor (called RFID sensor tag) is presented in the Fig. 1. It can be immediately noticed that some elements must be made in monolithic technology (e.g. RFID control system, memory, and also – in some solutions – sensors), but, in order to set up the assumed functionality, they need to be integrated with conventional thick-film structures (e.g. antenna, transducers of physical quantities into electrical signal, passive and matching elements, system of roots) [3, 4].

II. NAVIGATION SYSTEM

Aspects connected with the multi-functionality of electronic hybrid structures are presented on the example of a navigation system of static or moving objects within a given space – the system has been built on the basis of RFID technique with appropriate configuration of sensors and distributed memory of navigation data. Such system called ACRNM (Autonomous Robinson Crusoe Navigation Method) was proposed in [5] and its idea is presented in the Fig. 2.

In the ACRNM system, the RFID tags are placed according to a given structure on a plane or in space and serve as position sensors, and in addition, as sensors of other local environmental parameters (temperature, humidity, pressure, stored information about load-carrying capacity, surface roughness, gradient, dimensions of zone etc.). Such a defined network of sensors provides the necessary information to realize the location process of the moving object and its autonomous navigation without necessity of implementing any other intelligent detectors (e.g. cameras, ultrasonic and strain gauge, etc.).

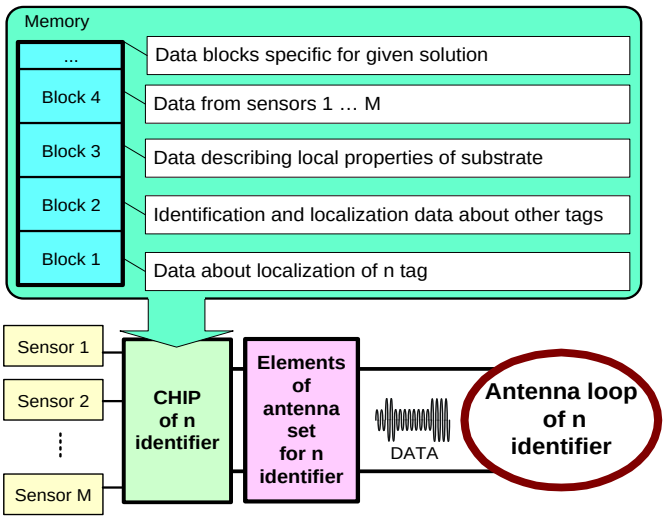


Figure 1. General idea of intelligent tag with sensors

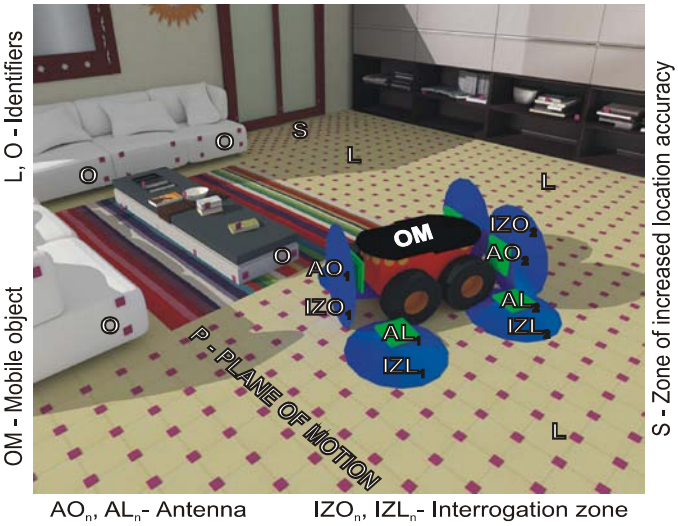


Figure 2. Idea of ACRNM navigation system

III. RFID IDENTIFIER – SENSOR IN NAVIGATION SYSTEM

The main application of RFID technique is the objects identification in which the identifier is the key element. The tag consists of electronic integrated circuit and antenna loop with inputs elements and it responds to inquiries from the reader / programmer (Read / Write Device, RWD). The primary information conveyed in the radio transmission is a unique serial number. This number can be unequivocally assigned to a particular grid node describing the moving area of object in the navigation system.

The simplest solutions of navigation systems with the possibility of informing mobile object about the environment characteristics can be constructed by using common identifiers equipped with data memory. Depending on the needs, the memory can be programmed in the design stage or modified in the process of message exchange with RWD. In this case, the stored data characterize local parameters which describe e. g. the type, slope and drop of substrate, sizes, shapes and properties of another object or fixed obstacles.

Using commercially available passive RFID tags equipped with a sensor of a physical quantity (there are tags with embedded temperature or humidity sensors, but also it is possible to find prototype solutions with other detectors) the significant increase in the visibility of environment properties is feasible. Moreover, the intelligence of navigation system is not necessarily more sophisticated.

The usage of a passive (powered by energy supplied by the antenna system) intelligent RFID platform with inputs for digital or analogue signals from e.g. sensor would seem the best solution for navigation system (like ACRNM). The idea of such solution is shown in Fig. 3. Feasibility of this type of tags is possible thanks to the emergence of the ICs (microprocessors and ASICs) with very low power consumption. Moreover, it was confirmed in laboratories of Intel Labs Seattle, where the so-called platform WSIP (Wireless Identification and Sensing Platform) was developed [6].

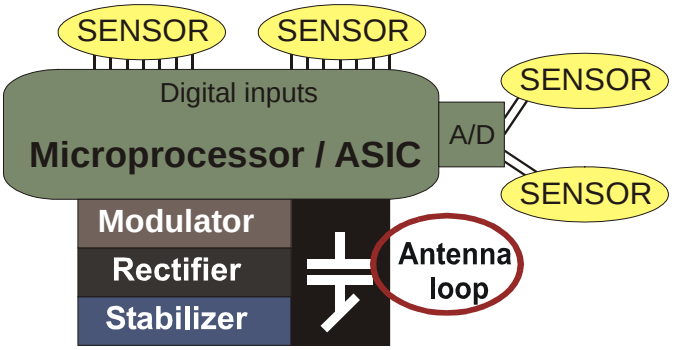


Figure 3. RFID tag platform for sensor - model under consideration

There is also an option to build general RFID platform by using commercially available transponders and additional digital switch (Fig. 4.). This device can receive and transmit signals from any sensor [7] by sampling the voltage value from the transducer and then activating selected antenna or whole block of identifier and broadcasting in real time a combination of different ID codes. Each combination of transmitted serial numbers corresponds to a different level of sensor output.

The similar solution could be made by using MCRF202 (Microchip Technology) tag which contains a 1-bit sensor port that can be used to invert the bit stream associated with the ID code and a Vcc port for powering external electronics (10µA/2V) [2].

Whatever solution someone would want to use there is always a problem with the rate of exchange data between IDs and RWD. The data transferring from the transponders in the network node takes place at the time when the tag is in the interrogation zone of the reader/programmer antenna localized on the mobile object. Moreover, a few identifiers may be present simultaneously in the interrogation zone and all off them may try automatically to answer to enquiry from the RWD. It causes the necessity of transmitting a huge number of bits connected either with serial number of identifier as well with anti-collision procedure. This process takes a significant amount of time and influences the speed of mobile object with RWD antenna [1]. Now, since there is the necessity to send extra bits from the built-in memory connected with data from sensors, the speed of the object has to be even more limited in consider network for navigation system. So, the problem is how to meet requirements for proper transmission and for intention of ambient data gathering.

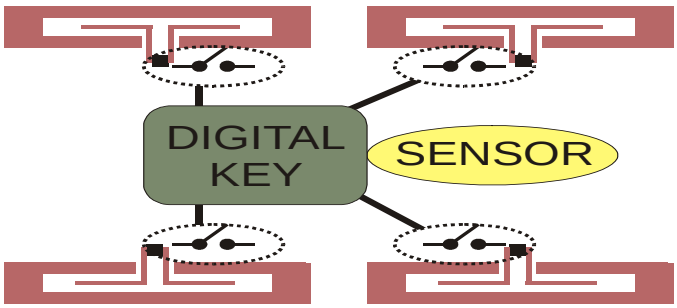


Figure 4. General purpose RFID tag platform for connecting generic sensors with 4-bits output

Since the identifier has its own unique serial number consisted of a few dozens of bits, the reading process of this identifying information is a significant part of the whole transmission. On the other hand, the position of given identifier with reference to others is sufficient to determine coordinates of autonomous object movement in the navigation system. So, there is possibility to code the relative position (for example according to eight neighbours) by using a few bits (for example one byte) which can replace more precise but long specific ID number. Of course this new dynamic code system has to ensure that all items in group of neighbours interrogated by RWD will have different allocated number. This approach greatly reduces the amount of transmitted data and increase ability of the mobile object to move at higher speed.

The memory organization of transponder for such solution is presented in the Fig. 5. The whole data which are designed for transmitting to RWD are divided into two main groups. The ”Basic information” is obligatorily transferred to control system every time. On the base of record “Status information” and operation algorithm the control system makes decision in what range the “Supplementary information” will be read. Of course the additional information can be completely omitted or read only partially. It means that the minimal number of data bits which are necessary to read from each identifier equals 10 bytes.

If there is requirement to make analysis of other setups it is possible to use following dependence:

$$LD(t) = \sum_{i=1}^{k} \left(10 + \sum_{j=1}^{8} status(j) * data(j) \right) \qquad (1)$$

where LD means current range of data designed to read, k is number of identifiers in interrogation zone and j means bits of “Status information” byte. If status(j) bit is true (Fig. 6), it means that there is possibility to read extra bytes data(j) which describe the network environment.

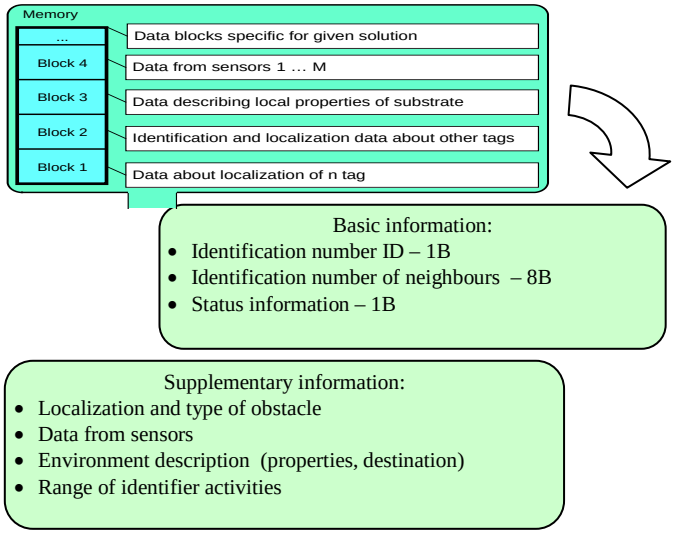


Figure 5. Memory organization of tags in navigation system of autonomous object

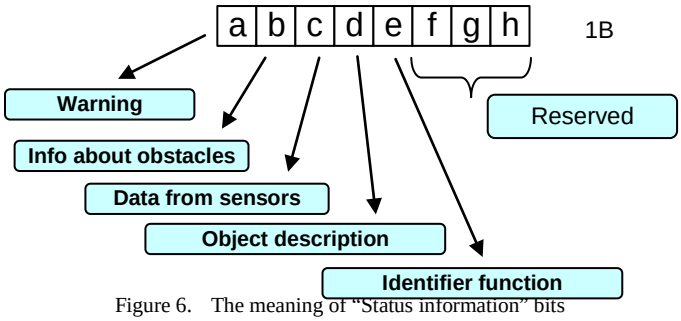


Figure 6. The meaning of “Status information” bits

The total number of transferred data is not constant and depends on the inquiries that are necessary to rout the path in navigation process realized by superior control system. This solution makes possible to increase the velocity of autonomous robot in not demanding environment.

IV. RFID SENSOR TAG IN ACRNM

The data transferring from the transponders in the network takes place at the time when the tags are in the interrogation zone of the reader/programmer antenna which is localized on the mobile object. Thus, important information for the current location of the object is delivered locally and at the time of passing the particular zone.

The two regular structures of identifiers distribution are presented in the Fig. 7 and Fig. 8. The arrangements (square and triangular) and trajectories of robot motion (Way 1, Way 2) have been selected on the base of previous works considered in other papers [8, 9, 10]. The numbers of identifiers being in the interrogation zone of the RWD antenna moving along the trajectories have been obtained on the base of numerical simulation (Fig. 9). The calculations have been made with assumptions that there is at least one identifier in the interrogation zone and maximal number of active nodes is four or three according to considered structure of lattice (square or triangle correspondingly).

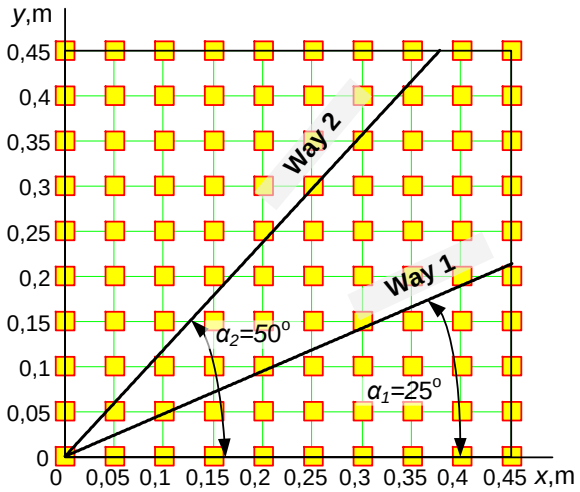


Figure 7. Considered trajectories of motions for square configuration of lattices

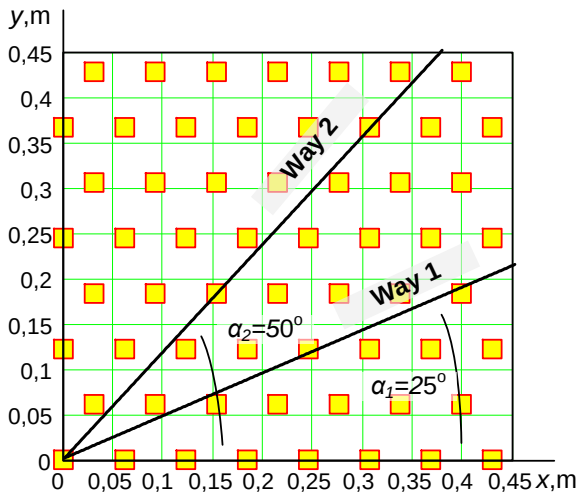


Figure 8. Considered trajectories of motions for triangular configuration of lattices

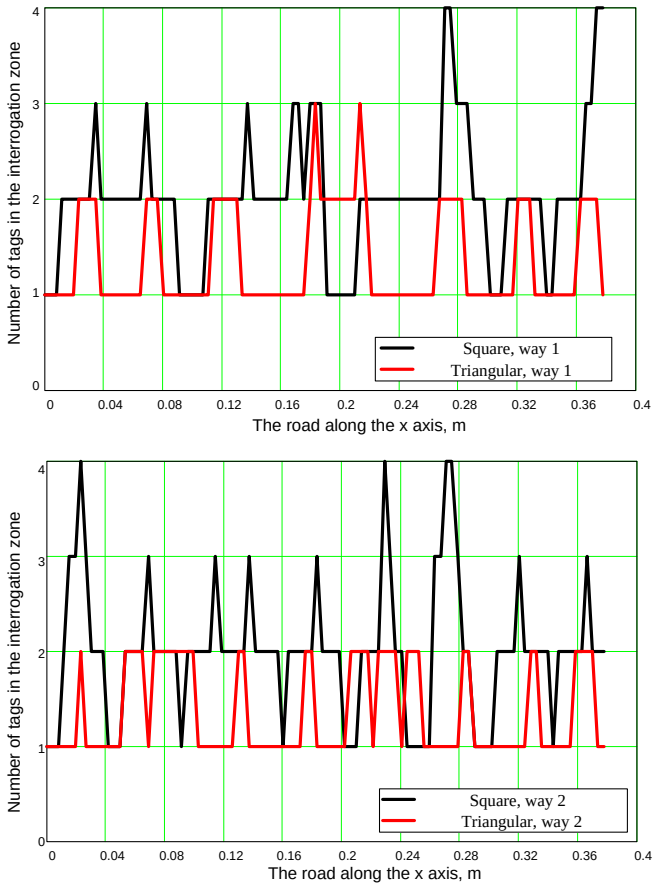


Figure 9. Number of RFID transponders in the interrogation zone for Way 1 / Way 2 and square / triangle latice

Trajectories have been chosen to take into account possible strong dynamic of changes and the fluctuation of the number of identifiers. These two parameters are important in terms of speed of the mobile object:

$$V = \Delta S_n / t_s \tag{2}$$

where ΔS_n means path along which the number of tags is constant in the interrogation zone (the direction of movement is consistent with one of n trajectories), t_s is duration of the data exchange with identifiers located in the interrogation zone.

The time needed to exchange data with identifiers currently situated in the interrogation zone is equal:

$$t_s = t_{AC} + \sum_{1}^{k} \{t_{Pk}, t_{READk}, t_{WRITEk}, \ldots\} \tag{3}$$

where $k \in [1, 2, 3, 4]$, t_{Pk} means duration of signal processing from sensor (sensors) and data writing in the memory of k–th identifier, t_{AC} – duration of handling non-collision procedure, t_{READk} – duration of data reading from k–th identifier, t_{WRITEk} – duration of data writing in k–th identifier.

t_{READk} and t_{WRITEk} depend on the amount of data exchange and rate – BitRate parameter [9].

V. SUMMARY

The usage of tags equipped with special systems of sensors gives a new quality in the process of building navigation systems. This allows both the location of the mobile object as well as the construction of cognitive knowledge base about the environment. Unfortunately, increasing the universality of markers (sensor tags) the necessity of reading vast amounts of data affects the speed of movement of the mobile object. In extreme cases $\Delta S_n \rightarrow 0$, that means too rapid change in the number of identifiers (or replacement of identifiers) in the interrogation zone. This situation leads to faults in processes of data transferring and thus – in an extreme cases – to losses in information from specific identifier (sensor). The maximum speed should be a compromise between the required rate of movement of the mobile object and assumed level of losses in data gathering.

ACKNOWLEDGMENT

The research was partially sponsored by the Polish National Science Centre (NCN), Project Grant No. 4711/B/T02/2011/40.

The research equipment was purchased in the projects:

- "Developing research infrastructure of Rzeszów University of Technology", No POPW.01.03.00-18-012/09 from The Structural Funds, The Development of Eastern Poland, The European Regional Development Fund;
- "Developing and modernization research base of Rzeszów University of Technology", No UDA-RPPK.01.03.00-18-003/10-00 from The Structural Funds, The Development of Podkarpacie Province, The European Regional Development Fund.

REFERENCES

[1] Y. Zhang, L.T. Yang and J Chen, RFID and sensors network, CRC Press, 2010

[2] D. Watters, P. Jayaweera, A. Bahr and D. Huestis: "Design And Performance Of Wireless Sensors For Structural Health Monitoring, Quantitative Nondestructive Evaluation", AIP Conf. Proc., Vol. 615, 2002, pp. 969-976.

[3] K. Finkenzeller, RFID Handbook: Fundamentals and Applications in Contactless Smart Card and Identification, SE, Wiley, 2003.

[4] D. Redinger, R. Farshchi and V. Subramanian, "An all-printed passive component technology for low-cost RFID", 33th IEEE European Solid-State Device Research Conference DIGEST'03, 2003, pp. 187- 188.

[5] W. Kalita and M. Skoczylas, "System nawigacji autonomicznych obiektów mobilnych z wykorzystaniem techniki RFID", Elektronika, nr 8, 2010.

[6] A. Sample, D. Yeager, P. Powledge and J. Smith, "Design of a Passively-Powered, Programmable Sensing Platform for UHF RFID Systems", IEEE International Conference on RFID, Grapevine, USA, 2007, pp. 149-156

[7] L. Catarinucci, R. Colella and L. Tarricone, "Sensor Data Transmission Through Passive RFID Tags to Feed Wireless Sensor Networks", Microwave Symp. Digest, 2010 IEEE MTT-S International 2010, pp. 1772-1775

[8] P. Jankowski-Mihułowicz, W. Kalita and B. Pawłowicz, "Problem of dynamic change of tags location in anticollision RFID systems", Microelectronics Reliability (2008), Vol. 48, Issue 6, pp. 911-918.

[9] P. Jankowski-Mihułowicz, W. Kalita, M. Skoczylas and M. Węglarski, "Wpływ struktury przestrzennego rozmieszczenia identyfikatorów-czujników RFID na jakość procesu sterowania autonomicznych obiektów", Elektronika, nr 6, 2010.

[10] P. Jankowski-Mihułowicz, W. Kalita and M. Skoczylas, "Identyfikator RFID jako sensor w systemie nawigacji autonomicznych obiektów", Elektronika, nr 6, 2010.

Security in Communications

chairman:

Michal HALAS

Femtocell Backhaul Security Efficiency

Matej Rohlik, Tomas Vanek
Department of Telecommunication Engineering
Czech Technical University in Prague
Technicka 2, Prague, 166 27 Czech Republic
matej.rohlik@fel.cvut.cz, tomas.vanek@fel.cvut.cz

Abstract—**One of the relatively new services presented by mobile operators is a femtocell. A femtocell is a network located at the residential premises helping to extend the mobile signal to places which are difficult to cover. Furthermore, it enables the mobile operator to provide attractive service to the customer since the femtocell is connected to the mobile operator network using an IP based backhaul link over the public Internet. To ensure appropriate security over the untrustworthy environment, an IPsec tunnel is established between the femtocell access point and the provider's security gateway located at the core network perimeter. IPsec itself wasn't originally proposed to carry small voice packets resulting in a redundant overhead. This paper examines other security procedures, such as transport layer security (TLS) and Datagram TLS (DTLS) protocols.**

Keywords-component; security, network, femtocell, protocol, IPsec, TLS, DTLS, SRTP, ZRTP

I. INTRODUCTION

A femtocell is a network located at the residential premises helping to extend the mobile signal to places which are difficult to cover. Furthermore, it enables the mobile operator to provide attractive service to the customer since the femtocell is connected to the mobile operator network using an IP based backhaul link over the public Internet. As the IP backbone Wide Area Network (WAN) topology, every available broadband technology (fiber, cable, digital subscriber line, etc.) technology can be used. The scenario is depicted in Fig. 1 [1].

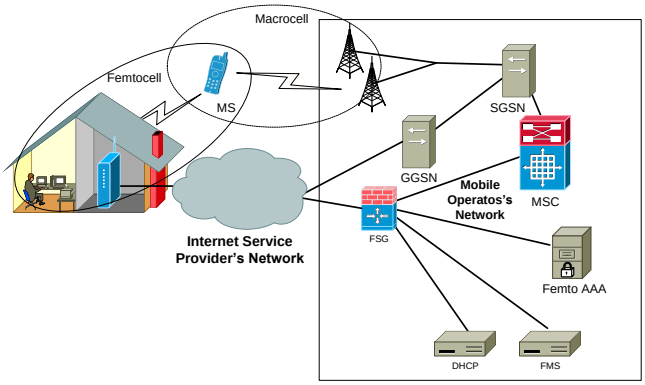


Figure 1. The Scenario of a Femtocell

The essential requirements imposed on the broadband technology are adequate transmission parameters such as jitter, delay, and bandwidth [2, 3].

II. IP BACKHAUL LINK FEMTOCELL SECURITY

A. *Architecture of the Femtocell*

The central appliance of the femtocell is a device named Femtocell Access Point (FAP). Such device acts as a common access point known from the WiFi networks. The difference is that the mobile station (MS) is connected to the FAP using the 2G, 3G, or pre-4G network (GSM/GPRS/EDGE, UMTS, LTE, WiMAX, HSDPA, HSUPA, etc.). The FAP is then connected to the Femto-Security Gateway (FSG), located and the mobile operator core network perimeter, through the public Internet.

Since the public Internet is an untrustworthy environment, the user data (voice or data packets) and the signaling (control packets) are encapsulated and encrypted in a previously established IPsec tunnel. The femtocells connected to the FSG are managed by the respective Femto Management System (FMS) and are authenticate and authorized against the appropriate Femto AAA server. This approach is depicted in Figure 2 [4,5].

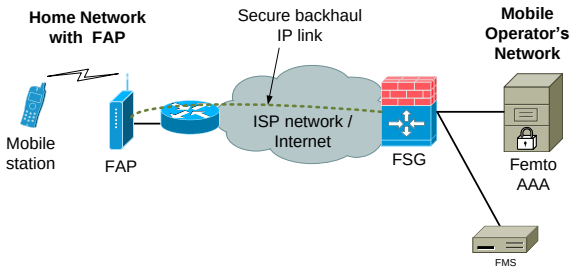


Figure 2. The Security Architecture of a Femtocell

The mutual authentication, between the network (FSG) and the FAP, is performed by the Internet Key Exchange version 2 (IKEv2) protocol with certificates [6, 7, 8]. The confidentiality

between the FAP and the FSG is provided by setting up the IPsec/ESP tunnel [9].

The traffic in the operator's network (between the FSG and the FMS) is secured as well since such link is considered insecure. However, the IPsec protocol was not designed to carry small voice packets and behind Network Address Translation (NAT) causes a relatively big packet overhead. We have examined other security approached to eliminate the IPsec disadvantages.

B. Considered Security Methods

This IPsec drawback can be eliminated using a Transport Layer Security (TLS) protocol defined in [10]. The TLS and SSL respectively, provide the end-points respective authentication and privacy. The practical approach, e.g., in WWW or email services, is to authenticate only the server and not the client. However, both, the server and the client can authenticate mutually. Similar to the IPsec protocol, the Public Key Infrastructure (PKI) keys can be applied to authenticate the communicating parties using certificates. TLS establishes an end-to-end secured session and is encapsulated into the reliable Transmission Control Protocol (TCP).

The TCP usage does not appear to be the optimal solution to carry the voice packets. Therefore, the Datagram TLS (DTLS) protocol was proposed to transport the data streams unreliably and is defined in [11]. The DTLS provides the same privacy to user data while using the User Datagram Protocol (UDP) as transport layer protocol.

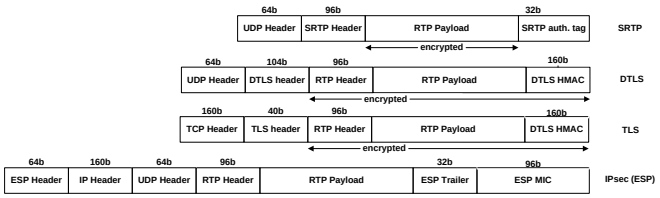


Figure 3. Selected security protocols overhead

For end-to-end data streams transport, Real-Time Transport Protocol (RTP) was designed. In conjunction with RTP Control Protocol (RCTP), RTP provides transport of the multimedia data and RTCP ensures Quality of Service (QoS) parameters [12]. Since none of the previously noticed protocols offer privacy and security to the carried data, another session oriented protocol was designed.

Secure RTP (SRTP) and Secure RTCP (SRTCP) were proposed to remove the disadvantage of the unsecured protocols [13]. Both utilize Advanced Encryption Standard (AES) as the default encryption algorithm in Segmented Integer Counter Mode. The message integrity is ensured by the HMAC-SHA-1 hash algorithm. As the extension of the original SRTP, ZRTP was developed [14]. ZRTP doesn't secure RTP streams itself, but uses a Diffie-Hellman protocol as a key-agreement method to negotiate the encryption keys for SRTP.

All of the previously noticed protocols, which provide respective security and privacy to the transported data, were taken into consideration as the adequate IPsec replacement with the main focus applied to the overhead, delay and jitter optimization.

III. RESULTS

The measurement simulated a voice over IP call with several different codecs, encapsulations and cryptography algorithms providing security. The application layer protocol considered as an encapsulation protocol to the voice stream was RTP. Assumed security mechanisms were no security protocol, SRTP, TLS, and DTLS. ZRTP was utilized to negotiate encryption keys for SRTP. As a transport layer protocol, UDP and TCP were used. The Simena Network Emulator NE2000 [15] was applied to simulate one of the basic parameters of network environments such as uplink/downlink bandwidth. The simulation scenario is depicted Figure 3.

The average length of the call was estimated to 1 minute. The measurements were accomplished in the range from 50 to 70 seconds since the measurement length does not affect the measurement itself.

Since FAP are assumed to be applied mainly in the household environment, an internet connection with parameters corresponding to a widespread technology ADSL (4 Mbps downlink and 0.25 Mbps uplink) was simulated. The home FAP is considered to operate in the "closed mode" where only allowed users (usually members of the family) are allowed to use the provided service. Therefore, only limited traffic can be considered (1 or 2 concurrent voice calls). As a consequence of the measured values, when QoS properly configured, low-utilization FAPs can be run at such low-speed connections.

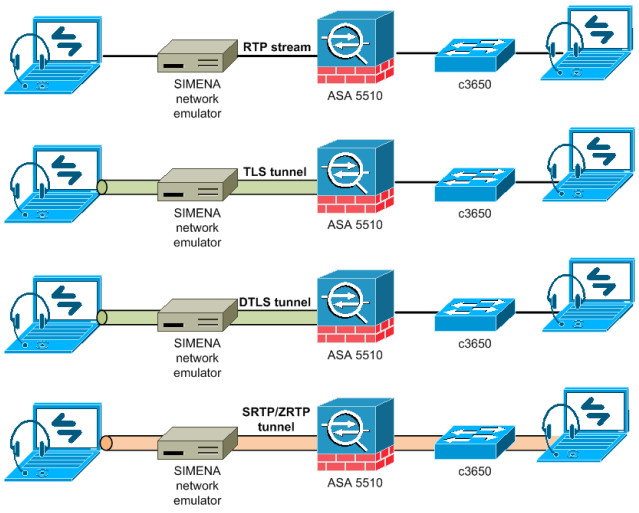


Figure 4. The scenarios of the Simulations

For different security methods simulation, we utilized a Cisco ASA 5510 appliance (a firewall) which enabled us to capture both secured and unsecured traffic and its ensuing analysis. In case where the voice streams were encrypted by the SRTP protocol, the call conversation was established directly between the communicating parties (end stations) and the stream was not affected by the ASA firewall. The encryption keys were negotiated using the ZRTP protocol.

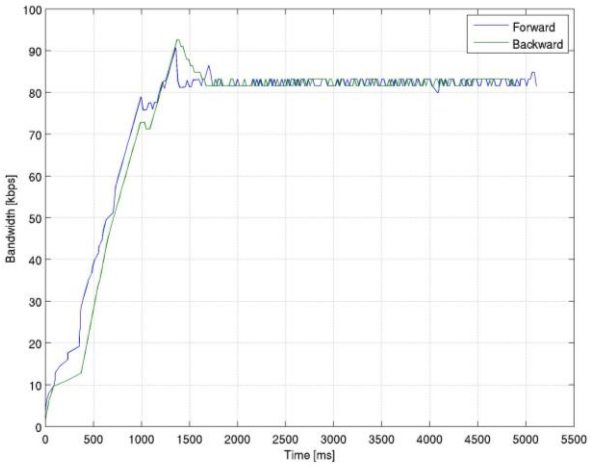


Figure 5. Bandwidth requirements for voice call with G.711 and SRTP

The Figure 5 depicts the results of the one of a number of measurements which were accomplished. The bandwidth requirement is shown for G.711 codec while RTP stream is secured via SRTP. The total bandwidth is higher than the values in the Table 1 because the graph illustrated in the Figure 1 includes the overhead of IPv4 protocol.

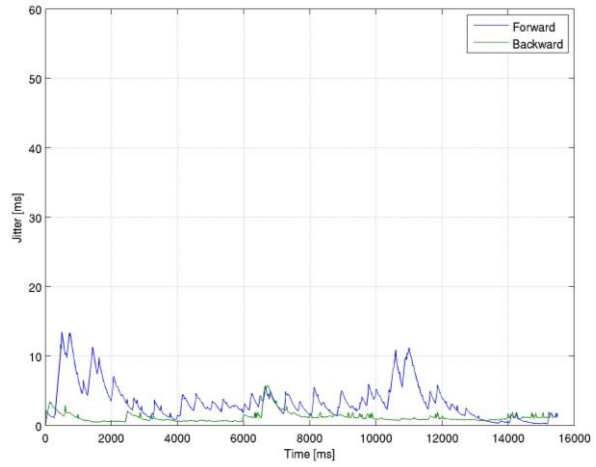


Figure 6. The jitter of SRTP secured voice call using G.711 codec

In the Figure 6, an example of measured voice call jitter is shown. The graph represents the asymmetrical character of the simulated line. The jitter is relatively higher in the uplink than in the downlink direction which is less loaded.

In the Table 1, the measured bandwidth requirements for a selected security protocol is shown. An RTP voice payload size was 20 ms for each codec. The mentioned bandwidths are related to the transport layer. The overall bandwidth requirements, when IPv4 based network layer and Ethernet link layer used, are increased by 60 B per packet, i.e., by 30 kbps.

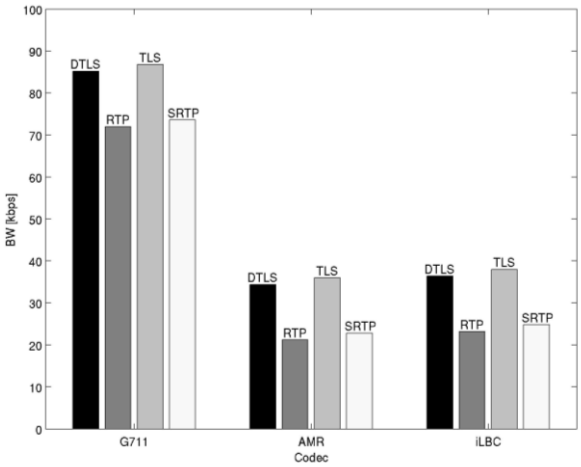


Figure 7. The scenarios of the Simulations

The Figure 7 shows the average values of the transmission rate for each protocol which were determined from the analysis of the measured values.

TABLE I. MEASURED THROUGHPUT ON TRANSPORT LAYER FOR SELECTED SECURITY PROTOCOLS. RTP VOICE PAYLOAD SIZE WAS 20 MS

Codec	**Codec bitrate [kbps]**	**RTP [kbps]**	**SRTP [kbps]**	**TLS [kbps]**	**DTLS [kbps]**	**IPsec (ESP in tunnel mode) [kbps]**
G.711	64.0	72.0	73.6	86.8	85.2	89.6
iLBC	15.0	23.2	24.8	38.0	36.4	40.8
AMR	13.2	21.2	22.8	36.0	34.4	38.8

TABLE II. MEASURED THROUGHPUT ON TRANSPORT LAYER FOR SELECTED SECURITY PROTOCOLS. RTP VOICE PAYLOAD SIZE WAS 10 MS

Codec	**codec bitrate [kbps]**	**RTP [kbps]**	**SRTP [kbps]**	**TLS [kbps]**	**DTLS [kbps]**	**IPsec (ESP in tunnel mode) [kbps]**
G.711	64.0	80.0	83.6	109.6	106.4	115.2
iLBC	15.0	31.2	34.8	60.8	57.6	66.4
AMR	13.2	29.6	33.2	59.2	56.0	64.8

Comparing the measured bandwidth values depicted in Tables 1 and 2, the usage of voice payload size of 10 ms enables the voice stream to have smoother pass over the network redeemed by a significant overhead growth. ZRTP increases the packet size by 4 B, i.e., for voice payload size of 20 ms by 1.6 kbps which is negligible and acceptable.

IV. CONCLUSION

Within the optimization of the femtocell backhaul link security methods, the existing cryptographic protocols (SRTP, DTLS, and TLS) were practically experimented and analyzed.

Even though the differences between the analyzed protocols are not large, the common location of a femtocell will be behind a NAT enabled device (ADSL modem, firewall, etc.)

In this case, the bandwidth requirements for IPsec increase for approximately 5 % because of the NAT-T mechanism. TLS and DTLS have similar bandwidth requirements, however, DTLS is based on UDP and such application is less susceptible to packet loss and out-of-order delivery.

Based on the analyzed results, as the most advanced and prospective solution to replace the complex IPsec mechanism is the combination of the TLS and DTLS protocols where DTLS will be applied as a voice stream security algorithm and TLS as signaling and FAP management security method. Within the next research, we would like to focus on the authentication process between FAP and FSG using the EAP-AKA protocol.

ACKNOWLEDGMENT

This research work was supported by MSMT under the project no. MSM 6840770038.

REFERENCES

[1] Chandrasekhar, V., Andrews, J. G., Gatherer, A., Femtocell Networks: A Survey. In IEEE Communications Magazine. 2008.

[2] Kim, R. Y., Kwak, J. S., Etemad, K., Wimax Femtocell: Requirements, Challenges, and Solutions. In IEEE Communications Magazine. 2009.

[3] Hasan, S. F., Siddique, N. H., Chakraborty, S., Femtocell Versus WiFi – A Survey And Comparison of Architecture and Performance. In Wireless Vitae'09. 2009.

[4] Security Issues in Femtocell Deployment [online]. 2008 [cit. 2011-03-28]. Available: http://www.gsmworld.com/documents/fcg0510.pdf.

[5] 3GPP2 S.S0132-0 "Femtocell Security Framework 1.0", http://www.3gpp2.org/public_html/specs/S.S0132-0_v1.0_Femtocell_Security_Framework.pdf

[6] Korver, B., The Internet IP Security PKI Profile of IKEv1/ISAKMP, IKEv2, and PKIX (RFC4945) [online]. 2007 [cit. 2011-03-28]. Available: http://tools.ietf.org/html/rfc4945.

[7] Kaufman, C., Internet Key Exchange (IKEv2) Protocol (RFC4306) [online]. 2005 [cit. 2011-03-28]. Available: http://tools.ietf.org/html/rfc4306.

[8] Black, D., McGrew, D., Using Authenticated Encryption Algorithms with the Encrypted Payload of the Internet Key Exchange version 2 (IKEv2) Protocol (RFC5282) [online]. 2008 [cit. 2011-03-28]. Available: http://tools.ietf.org/html/rfc5285.

[9] Manral, V., Cryptographic Algorithm Implementation Requirements for Encapsulating Security Payload (ESP) and Authentication Header (AH) (RFC4835) [online]. 2007 [cit. 2011-03-28]. Available: http://tools.ietf.org/html/rfc4835

[10] Dierks, T., Rescorla, E., The Transport Layer Security (TLS) Protocol Version 1.2 (RFC5246) [online]. 2008 [cit. 2011-03-28]. Available: http://tools.ietf.org/html/rfc5246

[11] Rescorla, E., Modadugu, N., Datagram Transport Layer Security (RFC4347) [online]. 2006 [cit. 2011-03-28]. Available: http://tools.ietf.org/html/rfc4347.

[12] Schulzrinne, H., Casner, S., Frederick, R., Jacobson, V., RTP: A Transport Protocol for Real-Time Applications (RFC3550) [online]. 2003 [cit. 2011-03-28]. Available: http://tools.ietf.org/html/rfc3550.

[13] Baugher, M., McGrew, D., Naslund, M., Carrara, E., Norrman, K., The Secure Real-time Transport Protocol (SRTP) (RFC3711) [online]. 2004 [cit. 2011-03-28]. Available: http://tools.ietf.org/html/rfc3711.

[14] Zimmermann, P., Callas, J., Johnston, A. ZRTP: Extensions to RTP for Diffie-Hellman Key Agreement for SRTPdraft-zimmermann-avt-zrtp-01. (RFC Draft) [online]. 2006 [cit. 2011-03-28]. Available: http://tools.ietf.org/html/draft-zimmermann-avt-zrtp-01.

[15] Simena Network Emulator NE2000, [online]. 2010 [cit. 2011-03-28]. Available: http://simena.net/products/network-emulator-2/

Optimizing brute-force attack on MD5

Dalibor Hula
Department of Informatics, School of Business Administration in Karviná
Silesiaian University in Opava
Karviná, Czech Republic
hula@opf.slu.cz

***Abstract*—This paper describes various ways how the brute-force attack on MD5 algorithm can be optimalized. Text was prepared as a part of the project SGS/24/2010.**

Keywords – MD5, hash function, optimization

I. INTRODUCTION (HEADING 1)

All companies and organizations which are providing some kind of online services with authentication must deal with security. One of the most popular hash algorithms that are widely used to store passwords in the database or to authenticate users is MD5. This old algorithm has many shortcomings that can be used by attackers to optimize their attack.

SIP is an example of protocol which in default utilizes MD5 hashing function to authenticate user. Retrieving user login information from MD5 hash value is quiet fast with modern hardware.

There are a lot of papers about finding collisions in MD5 and MD5 hash function is now considered as not collision-resistant, or simply "broken".

II. MD5 ALGORITHM

MD5 is cryptographic hash function designed in 1991 by Ron Rivest and specified in RFC 1321 [1]. It is a successor of MD4 hash function, which didn't satisfied the security needs. Although MD5 is also considered as "broken" and not collision resistant, it is still widely used to check the integrity of files or to authenticate user passwords.

MD5 produces 128-bit hash value output from the input message (length of the message is not limited). An MD5 hash is typically expressed as a 32-digit hexadecimal number.

The input string is divided into parts with a length of 512 bits, which represent sixteen 32-bit numbers. In the last part, binary 1 is appended to the end of the message followed by binary 0's up to 448th bit. Last 64 bits contains integer representing the length of the original message in bits.

MD5 hash calculation itself uses four 32-bit variables a, b, c, d, which are initialized at the beginning by constant values. The algorithm processes separately the 512-bit parts in four similar phases (rounds). Each individual round is composed of 16 similar steps, where each step consists of several logical

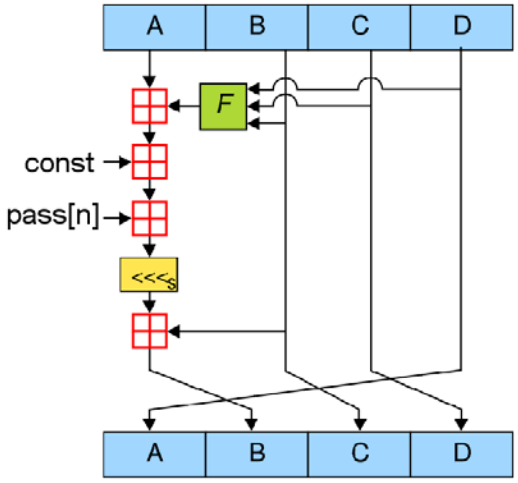


Figure 1: One step of MD5 function. Value "const" is unique for each one of the 64 steps, pass[n] denotes 32-bit part of the original message. Source: Wikipedia.org

bit operations, additions and bit rotation as shown on Fig.1

Every single step in MD5 algorithm has the form:

$$a = b + ((a + f(b, c, d) + data + const) <<< const)$$

where <<< denotes left bit rotation

f is nonlinear function

Function f is different for each round as follows:

Round 1: $f_1(b, c, d) = (b \wedge c) \vee (\neg b \wedge d)$

Round 2: $f_2(b, c, d) = (b \wedge d) \vee (c \wedge \neg d)$

Round 3: $f_3(b, c, d) = b \oplus c \oplus d$

Round 4: $f_4(b, c, d) = c \oplus (b \vee \neg d)$

where symbols $\wedge \vee \oplus \neg$ stand for logical bit operations AND, OR, XOR and NOT.

III. OPTIMIZATIONS

A. Reversing hash

The base optimization practice is to pre-compute as much as possible. This paper focuses on brute-force attacks, so techniques like "Rainbow tables" or pre-image attacks will not be discussed. Many MD5 steps can be pre-computed.

As mentioned above, the original message (password) is padded by zeroes and divided into fourteen 32-bit elements. Together with 64 bits representing the length it forms an array pass[] with sixteen 32-bit values.

Let's see what elements of the array pass[] are gradually used at different steps of MD5. The following list indicates indexes of array pass[] as they are processed.

Round 1: 00,01,02,03,04,05,06,07,08,09,10,11,12,13,14,15

Round 2: 01,06,11,00,05,10,15,04,09,14,03,08,13,02,07,12

Round 3: 05,08,11,14,01,04,07,10,13,00,03,06,09,12,15,02

Round 4: 00,07,14,05,12,03,10,01,08,15,06,13,04,11,02,09

Element pass[14] contains the length of the password, so it can be treated as a constant.

Element pass[15] will be zero, unless the length of password exceeds 2^{32} bits.

If we consider only passwords with length < 16, then all elements from pass[04] to pass[13] will be equal to zero. So pass[00] – pass[03] are only elements which will be changed during brute force attack.

The final step of the whole MD5 looks like this:

```
b = b + f(c,d,a) + const + pass[09];
rotate_left(b, const);
b = b + c;
```

From the final hash which we want to crack, we know values of variables a, b, c and d. As pass[09] is zero, we know values of all variables, so we are able to perform inverse computation and find value of variable b before this last step. The whole step is reversible:

```
b = b – c;
rotate_right(b, const);
b = b – f(c,d,a) – const – pass[09];
```

To fully exploit this possibility of inverse computation, password candidates must be wisely generated. At first, we will be changing only first 4 characters of password. Thus, for example if we are looking for password of 5 characters, we will be generating sequences like "aaaaa", "aaaba", "aaaca", … , "aaaza", "aaaab", "aaabb", "aaacb", … "aaazb",etc. If we are permuting only first 4 characters, only element pass[00] will be changed, elements pass[01], pass[02] and pass[03] will remain the same. In this case it is possible to revert all steps up to 49th step where previously unknown pass[00] is used. Moreover, it is possible to partially reverse next two steps. Most of the time, 43rd step is the last full step needed to compute to check if the correct password is found.

This process significantly reduces time to crack MD5 hash because only 67% (43/64) of the full MD5 algorithm is needed. Many programs implemented this technique, for example BarsWF, EmDebr, Hashkill, etc.

Every time we have to "re-reverse" (due to change in pass[01], pass[02] or pass[03]), we can also pre-compute constant additions in steps using pass[01], pass[02], pass[03].

B. More pre-computation

There is also possibility to pre-compute the whole first and a half of the second step. First two steps:

```
a = pass[00] + const;
a = ((a<<const) | (a>>(32-const)) ) + const;
d = pass[01] + f(c, d, a) + const;
d = ((d<<const) | (d>>(32-const)) ) + const;
```

Most of the user passwords is composed only from a limited character set. Let presume that character set includes only characters 'a'-'z', 'A'-'Z' and '0'-'9'. This means 62 possible characters. The number of possible combinations for element pass[00] is then 62^4 = 14 776 336 combinations. Each combination consist of 4 characters, so if we want to store them, we will need 59 105 344 bytes of memory. Nowadays, 56MB is not so much and we are able to pre-compute and store variable "a" for all of the combinations. Expression f (c,d,a) + const should be pre-computed too, but additionally 56 MB will be required.

This simple idea results in 2-3% speedup.

C. 3rd round optimization

Let's take a look at the first two steps in 3rd round of MD5 in C-like syntax:

```
a += ((b ^ c) ^ d) + const + pass;
a = ((a<<const) | (a>>(32-const)) ) + b;
d += ((a ^ b) ^ c) + const + pass;
d = ((d<<const) | (d>>(32-const)) ) + a;
```

We can see that operations (b ^ c) ^ d and (a ^ b) ^ c are performed. Because XOR operation is associative, we can write (a ^ b) ^ c = a ^ (b ^ c). Value (b ^ c) is common to both expressions, so it can be stored after the first step and directly used in the second step:

```
tmp = b ^ c;
a += (tmp ^ d) + const + pass;
a = ((a<<const) | (a>>(32-const)) ) + b;
d += (tmp ^ a) + const + pass;
d = ((d<<const) | (d>>(32-const)) ) + a;
```

All pair steps in round 3 can be rewritten in that manner, i.e. computing tmp in each even step and using it in each odd step. Round 3 consists of 16 steps, so we are able to avoid

8 XOR instructions. The result of this optimization depends on the architecture. On newer ATI GPUs (line 5xxx), it brings up to 2,2% speedup [2].

This small improvement can be applied also to regular MD5 or MD4 computation, Simply modified x86-64 code for MD5 from OpenSSL library outperforms original implementation by approximately 0,5%

D. Rotation optimization

Each step in MD5 consists of 1 bit rotation operation – left bit rotation. Assuming that in one step there is 4 additions, 2 - 3 bit operations and 1 bit rotation, these rotations have quite significant effect on overall performance.

Not all architectures (instruction sets) provide an easy way to perform bit rotation. SSE2 or nVidia CUDA are such examples. The most common way how to perform left bit rotation by n bits on 32-bit variable x is to compute:

```
x = (x<<n) | (x>>n)
```

where << and >> denotes bit shift operations which are more common operations (and could be easily implemented with multiplication and division when needed)

Using SSE2 instruction set, the above expression will be translated into 4 instructions. An extra register to store temporary value is needed:

```
MOVDQA tmp,x
PSLLD tmp, n
PSRLD x, n
POR x, tmp
```

If we look at the constants (n) which are used during MD5 computation, we will find that rotation by 16 bits is performed in steps 35, 39, 43 and 47. In this case, instead of rotating 32-bit value by 16 bit, we can only shuffle lower and upper 16 bits of that value. There are a lot of shuffling instructions in SSE2 instruction set. For our purposes, instead of using above 4 instruction code, we can use only two SSE2 instructions:

```
PSHUFHW x, x, 0xB1
PSHUFLW x, x, 0xB1
```

It means we can save 2 instructions and one temporary register. This minor optimization of SSE2 code results in about 2% better performance.

IV. FUTURE WORK

There are much more ways to improve brute-force attack on MD5. One of the possibilities which should be considered is “bitslicing” method (especially in combination with FPGA hardware).

Some minor optimization could be also done via exploiting some constant values. For example, in 18th step of MD5, constant value 0xc040b340 is added and the result value is rotated. Last six bits of this constant are zeroes, so after addition and rotation, those 6 bits won’t change allowing us to partially avoid these two operations (utilizing “meet in the middle” principle).

V. CONCLUSION

Brute-force attact on MD5 can be significantly optimized by various methods. According to these facts, using of MD5 algorithm is not recommended in the corporate sector. SHA1 or others “stronger” algorithms are better choice for these purposes.

REFERENCES

[1] R. Rivest, “The MD5 Message Digest Algorithm”, RFC1321, Internet Activities Board, 1992.

[2] I. Golubev, Www.golubev.com [online]. 22.11.2010. “GPU speed estimations for MD5/SHA1/Office 2007/WPA/WinZip”. WWW: <http://golubev.com/gpuest.htm>.

[3] J. Black, M. Cochran, C. A Highland, “Study of the MD5 Attacks: Insights and Improvements.” FSE 2006. WWW: <http://www.cs.colorado.edu/~jrblack/papers/md5e-full.pdf>.

[4] D. Hula, “MD5 algorithm from the attacker point of view” Textbook of the International Conference MMK 2010. ISBN 978-80-86703-41-1.

Finding the Best Encryption Algorithms for PLC Technology

Milos Orgon
Dept. of Telecommunications, FEI STU,
Ilkovičova 3, 812 19 Bratislava, Slovak Republic,
orgon@ktl.elf.stuba.sk

Michal Halas
Dept. of Telecommunications, FEI STU,
Ilkovičova 3, 812 19 Bratislava, Slovak Republic,
halas@ktl.elf.stuba.sk

Ivan Bestak
Dept. of Telecommunications, FEI STU,
Ilkovičova 3, 812 19 Bratislava, Slovak Republic,
bestak@ktl.elf.stuba.sk

Adrian Kovac
Dept. of Telecommunications, FEI STU,
Ilkovičova 3, 812 19 Bratislava, Slovak Republic,
kovaca@ktl.elf.stuba.sk

Abstract— **Power line communication (PLC) has become an important network access technology to conveniently and effectively provide various useful services such as remote meter reading, broadband Internet access, and home networking services. As the deployment and use of PLC networks increase, it is essential that we need to be able to configure, monitor, and control the PLC network resources. Results of this article shows the need for finding the most suitable encryption algorithm for use in PLC equipment and procedure for measuring performance of encryption algorithms. Our choice of the most suitable encryption algorithm can influence the transfer rate on the transmission route.**

Keywords-component; Power line, encryption, decryption, performance measurement, encryption algorithms

I. INTRODUCTION (HEADING 1)

Power Line Communication (PLC) has received tremendous attention in recent years as an alternative and cost-effective last-mile-access technology [1, 2]. It is easy to install PLC networks cost-effectively because the power lines already exist everywhere. Thus, power lines are frequently regarded as the preferred medium for providing a broadband connection to rural or remote areas where telephone and cable connections do not exist. PLC can provide data services, such as broadband Internet service, VoIP service, and a variety of value-added services, such as remote metering, street light control, home security, home appliance control, and many more. Efforts are also underway to provide the monitoring and control capabilities of their own power line equipment and facilities. A number of PLC systems have been installed worldwide for various kinds of field trials to demonstrate its feasibility and effectiveness (Fig. 1) [3, 4, 5].

II. PLC NETWORK SECURITY

Power Line Communications (PLC) is the name given to the technology which passes information (as an ADSL signal) from narrow band to broadband over an electrical circuit. This cabled electrical network may be in a house or a flat, for example. PLC works by superimposing a modulated signal at a higher frequency than the 50 Hertz electrical current delivered by your energy supplier to power your domestic electrical equipment.

Figure 1. The use of PLC systems

The transmission of data over Power Line is safer than most communications methods. It is easier to intrude on information being sent through commonly wired communication media. The security model combines up to five of the following protection methods:

- User Authentication and Authorization: Since the access system is Master-Slave based, any Slave node that wishes to connect to the network, and thus have access, must be recognized and authorized by the network Master. This authorization process is done with encrypted communication and is therefore considered secure. Any Slave node that has not been explicitly authorized to connect to a network will not be assigned any resources on the network, and all transmissions it may attempt will be disregarded.
- Encryption Method: PLC modem offers a combination of 3DES and DES encryption. The procedure ensures very strong security by using 3DES encryption for long-term information (encryption key) and a short key life (one single data burst) for DES encryption, thus keeping hardware costs low.
- Real-time Variable User-specific Modulation: Any message that is transmitted on a PLC network is modulated according to the SNR measured at each connection, and this is changed in real time as line characteristics change. In order to demodulate a message addressed from one user to another, the exact modulation used must be known, which means that an eavesdropper would have to intercept both upstream and downstream communications. In order to do this, the eavesdropper would need an SNR in the entire frequency band equal to, or better than, the equipment from which he or she is trying to intercept. In addition, the modulation information is encrypted, thus making eaves dropping a very difficult task.
- PLC modem technology also supports 802.1Q VLAN, which effectively isolates all users from each other and is completely controlled by the gateway, which is owned and controlled by the network operator. Access to management features of PLC devices can only be gained by means of a restricted access VLAN that can only be used by the network operator.
- Encrypted VPN is available for use on PLC networks.

Used DES / 3DES Encryption

The 3DES encryption/decryption block embedded in PLC equipment provides total isolation between networks and guarantees the privacy of network users. It consists of mixed DES/3DES encryption, configurable per PLC port and PLC node. Each connection on the network has a different 56-bit or168-bit key and encryption can be independently enabled/disabled. The key length is defined at the time of firmware compilation. A key length of 168 bits is only available with firmware that is compiled to support it.

3DES encryption is obtained by passing the data block through the DES cipher block three times, using three different 56-bit keys. International cryptanalysts consider DES and its variant 3DES high security algorithms that can be only broken using exhaustive search methods, also known as brute-force attacks.

Encryption algorithms and their use are essential part of the secure transmission of information. Wide range of encryption algorithms and their scalability allows you to select the appropriate algorithm for the area of deployment. Choosing the wrong algorithm may lead to security vulnerabilities of transmission routes but also to a significant delay in data transmission.

Encryption algorithms can be defined on the basis of several characteristics, including e.g. resistance to known attacks, the length of input block, length of encryption key and its scaling, power and age of the encryption algorithm [6]. As we will see in the following results, the time required for encryption and decryption, also belongs to parameters that significantly affect the selection of a suitable encryption algorithm.

III. TESTING SECURITY IN COMMUNICATION NETWORKS PLC

For practical testing of the security communication networks in PLC adapters we used PLC type Corinex AV200 used by Spanish company DS2 chipsets. When connecting to the adapters, which have a predefined IP address.

Adapters were connected to a network of computers after setting of adapters. Within this network we have brought in another adapter of PLC, which already was set on Network2 and other network encryption key. This engagement is shown in Fig. 2.

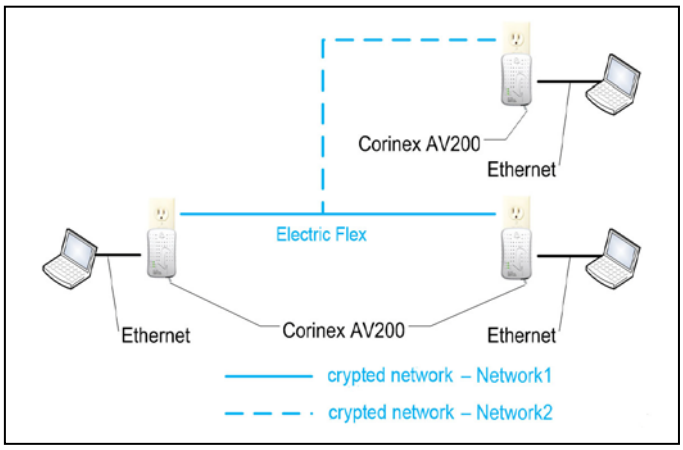


Figure 2. Involvement of two PLC adapters using one adapter with encrypted traffic from another network

IV. ENCRYPTION IN PLC NETWORKS

Encryption algorithms can be defined on the basis of several characteristics, including e.g. resistance to known attacks, the length of input block, length of encryption key and its scaling, power and age of the encryption algorithm [9]. As

we will see in the following results, the time required for encryption and decryption, also belongs to parameters that significantly affect the selection of a suitable encryption algorithm.

We decided to use the selected algorithms listed in Table I. of Crypto++ libraries, which are frequently used and implemented in existing systems and networks for measuring the time required for encryption and decryption and subsequent determination of the speed of encryption algorithms and their comparison. Among the selected algorithms, we also include Rijndael and 3DES [7], which are considered most commonly used in the PLC devices.

TABLE I. ENCRYPTION ALGORITHMS SELECTED FROM THE CRYPTO++ LIBRARY

Cipher	Block length [bit]	Key length [bit]
Rijndael AES	128	256
Serpent	128	256
Twofish	128	256
Cast6 (256)	128	256
Gost	64	256
Xtea	64	128
Blowfish	64	128
Cast5 (128)	64	128
3DES	64	168

Used hardware resources

Important role for testing the performance of encryption algorithm plays hardware and software computer equipment. The current notebook with the parameters listed in Table II**.** has been deliberately used in our test to determine the results achieved by conventional hardware and software resources.

TABLE II. USED HARDWARE RESOURCES

Type:	Notebook Toshiba Satellite A200
Motherboard:	Intel PM965 (Crestline-PM) + ICH8M
Processor:	Intel Core 2 Duo T7100 (1800 MHz)
Memory:	2048 MB DDR2-SDRAM
Graphics card:	ATI Mobility RADEON HD 2600
Graphics card memory:	256 MB DDR2 SDRAM 128-bit
Operating system:	Windows Vista Home Premium 32-bit
Software used:	Microsoft Visual C++ 2008

V. PERFORMANCE TEST OF CRYPTOGRAPHIC ALGORITHMS

The actual test of performance of encryption algorithms was carried out according to [3] and implemented in the environment of Microsoft Visual C++. The test consisted of measuring the time required for encryption, plus the time needed to decryption, since we need to execute both operations on the transmission path from source point to destination point. Size of input data ranged from 2^{n+4} with the value n = 0,1,2,3,..., 23 which represents the value from 16 bytes to 128 megabytes. To increase the accuracy of the measurement operation of encryption and decryption was performed several times depending on the size of input data, the number of repetitions of operations is doubled if not reached the minimum measurement time, which we experimentally chosen to 0.5 seconds. The initial number of repetitions depends on the value of 64 kilobytes in size as 16 bytes of input data represents a value of 4096 replicates. Vice versa, the minimum value of repeating the process of encryption and decryption, we chose to be 4, because since approximately 8192 bytes we can observe linear increase in time, therefore high accuracy is no longer necessary as it was for small values of input data when encryption algorithms are initialized respectively ”warmed up.". Repetition of whole cycle of measurement is performed 16 times for each encryption algorithm.

At Fig. 3 is flowchart which shows how measurements of the time required for encryption and decryption that, in dependence on the size of the input data for each encryption algorithm were carried out.

VI. MEASUREMENT RESULTS

The measurement result for each input data value is the average time in seconds calculated by (1) needed for encryption and decryption of all replicates and also all sixteen cycles of measurement. A corrected result is then calculated from the previous result by calculating the sample variance by the relation (2). From sample variation is then calculated standard deviation by relation (3). Also, the measurement results includes encryption speed and decryption speed in MB/s and also the maximum and minimum time required for encryption and decryption, depending on the size of the input data and subsequently calculated maximum and minimum speed MB/s.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad (1)$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right) \qquad (2)$$

$$s = \sqrt{s^2} \qquad (3)$$

These results are graphically illustrated in the figures below, where we can clearly see the impact of encryption and decryption on data transfer.

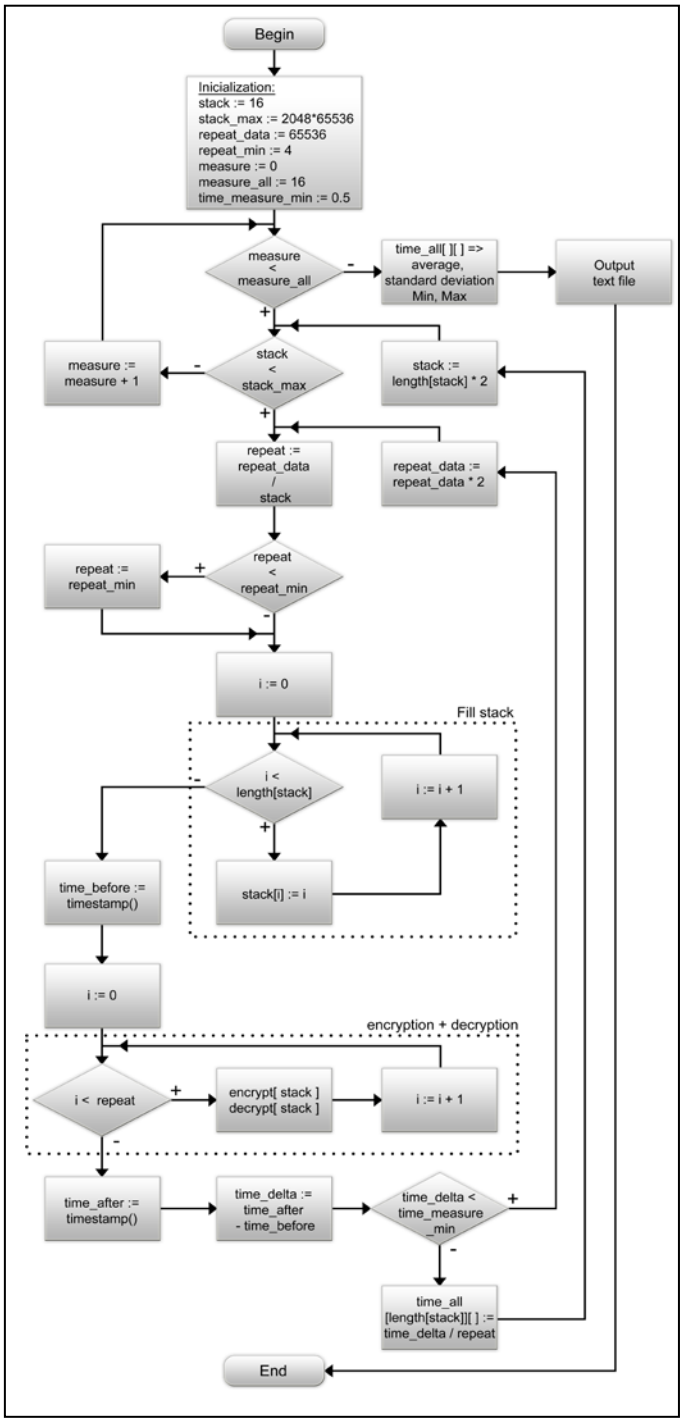


Figure 3. Flowchart for performance measurement

Figure 4 shows the duration of the encryption and decryption operations depending on the size of input data, which are in the range from 16 to 8192 bytes, to highlight the importance of the initial phase of the encryption and decryption. This phase is called to initialize or "warming up" of encryption algorithm and it is caused by planning of tables of encryption keys, preliminary calculations, the overhead of libraries and memory allocation.

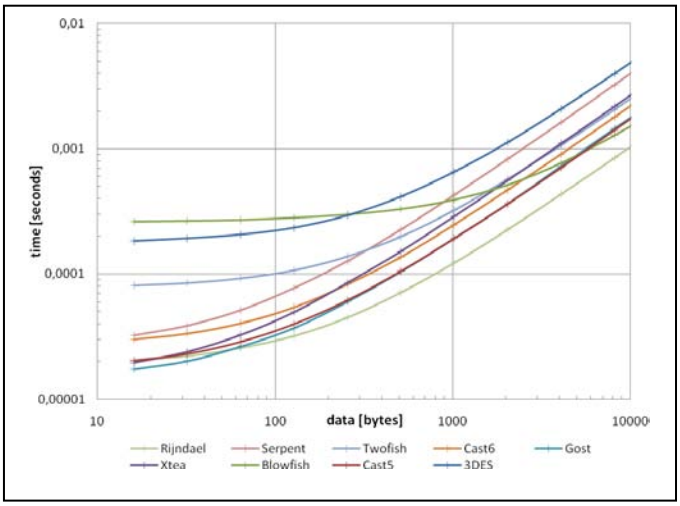


Figure 4. Dependence of time in seconds needed for encryption and decryption of input data size (16 to 8192 bytes)

As we can see at Fig. 4 algorithm Gost, CAST5, Xtea and Rijndael achieved the best results with the initialization, thus the time needed to encrypt and decrypt was the shortest, while the algorithms with the worst initial phase includes 3DES, Blowfish and Towfish. Encryption and decryption of 16 bytes input block with 3DES algorithm was about the same as the time required for Rijndael to encrypt and decrypt 2048 bytes which is about 128 times more encrypted and decrypted input data in the same time interval.

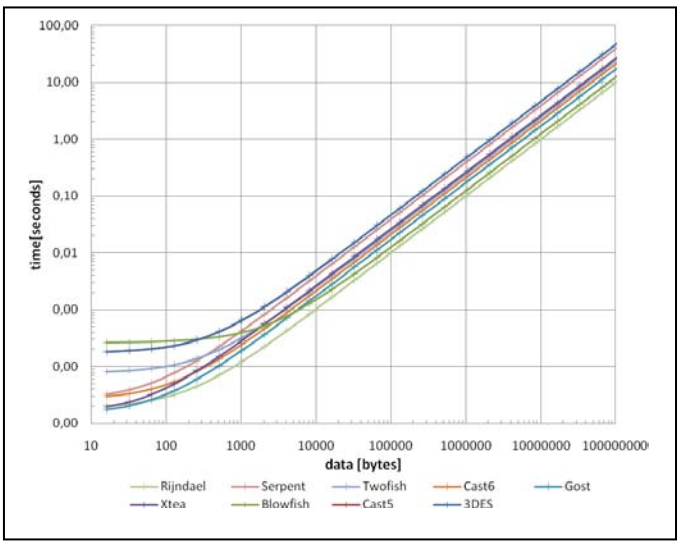


Figure 5. Dependence of time in seconds needed to encrypt and decrypt the size of the input data (16 to 134217728 bytes)

As we can see at Fig. 5 after the initial initialization, which has no effect on the value of approximately 8192 bytes, the time required for encryption and decryption increases linearly at all algorithms, depending on the size of the input data. Interesting is that Blowfish algorithm, after the slowest initialization phase, belongs together with the Rijndael algorithm between algorithms with the shortest time required

for encryption and decryption. 3DES algorithm has always needed most time with the same size of input data.

The next section displays a graphical dependency between encryption speed of algorithms and the size of input data. Encryption speed was calculated by dividing the size of the input data in to the average time needed for their encryption and decryption. Based on this dependency, we see how will the resulting bandwidth for the transmission path between two end points change if we use data encryption algorithm, or we transmit data in the open form.

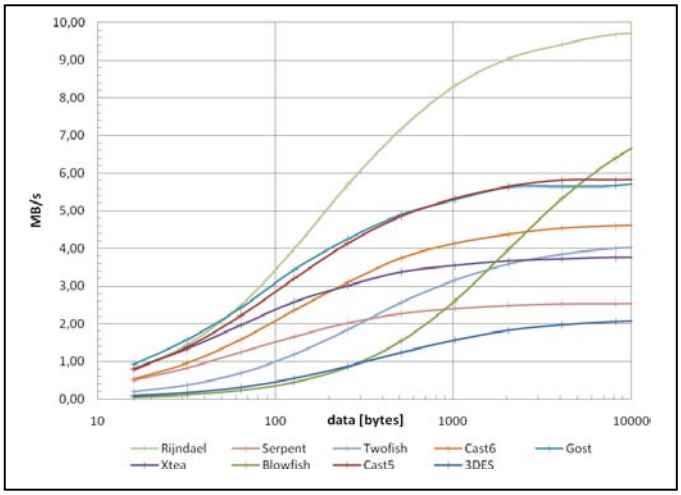


Figure 6. Example of a figure caption. (figure caption)

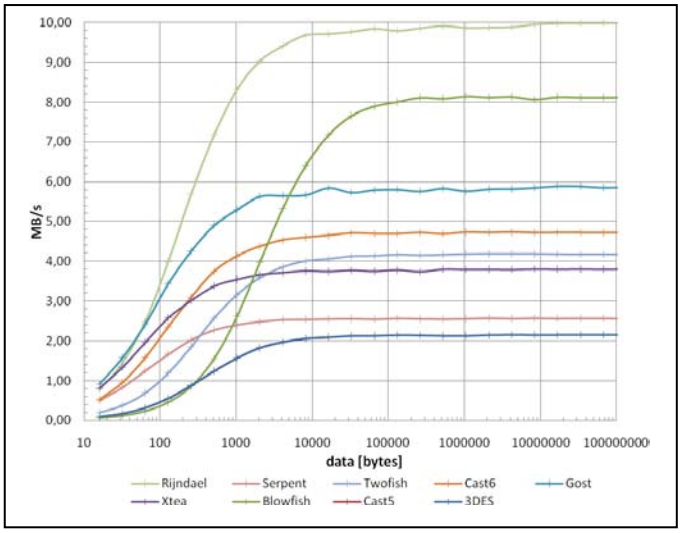


Figure 7. Example of a figure caption. (figure caption)

On figure 6 we can see that the encryption algorithm Rijndael already in 8192 Bytes gets to its maximum speed of encryption and approximately 10 MB/s, which is approximately constant from this size of the input data. Also, Blowfish, who after a very bad start initialization phase, which lasts up to a value of 10 megabytes, is the second fastest algorithm in our test with a maximum reached speed about 8 MB/s.

Currently very often implemented encryption algorithm 3DES is the slowest algorithm in our test. In comparison with the algorithm Rijndael, 3DES algorithm moves only about 20% of its power. The maximum speed achieved, as we can see at figure 7 moves just above 2 MB/s which is for the current systems in terms of performance, considered insufficient.

VII. Conclusion

Securing communication by encryption algorithms significantly limits the speed at which data are transferred. Communication speed between endpoints is reduced to a rate at which endpoints are capable of encrypting and decrypting the input data and thus the actual speed at which devices are able to communicate is limited. It is the choice of the right encryption algorithm that allows you to eliminate this influence to a certain extent. Therefore choice about using an encrypting algorithm should be based on a real need to secure transferred information.

Acknowledgment

This work is a part of research activities conducted at Slovak University of Technology Bratislava, Faculty of Electrical Engineering and Information Technology, Department of Telecommunications, within the scope of the ITMS 26240120029 “Support for Building of Centre of Excellence for SMART technologies, systems and services II”.

[1] Pavlidou N, Han Vinck AJ, Yazdani J. Power line communications: state of the art and future trends. *IEEE Communications Magazine* 2003; April: 34–40.

[2] . Majumder A, Caffery J. Power line communications. *IEEE Potentials* 2004; 23(4): 4–8.

[3] Jee G, Rao RD, Cern Y. Demonstration of the technical viability of PLC systems on medium- and low-voltage lines in the United States. *IEEE Communications Magazine* 2003; 41(5): 108–112.

[4] Liu W, Widmer H, Raffin P. Broadband PLC access systems and field deployment in European power line networks,.*IEEE Communications Magazine* 2003; 41(5): 114–118.

[5] Mlýnek, P.; Mišurec, J.: Systémy PLC pro dálkový sběr měřených dat. *Elektrorevue - Internetový časopis (http://www.elektrorevue.cz),* 2009, roč. 2009, č. 1, s. 1-5. ISSN: 1213- 1539.

[6] Piper, F.,Murphy, S.: Kryptografia – Průvodce pro každého. Praha: Dokořán, 2009, ISBN 80-7363-074-5.

[7] Levický, D.: Kryptografia v informačnej bezpečnosti (Cryptography in the Information Security), Košice: elfa, 2005, ISBN 80-8086-022-X.

[8] Speedtest and Comparsion of Open-Source Cryptography Libraries and Compiler Flags. [online] 2008, http://idlebox.net/2008/0714-cryptography-speedtest-comparison/

[9] Bešťák, I., Orgoň M.: Bezpečnosť komunikácie v sieťach elektrických rozvodov (Security of communication in electrical distribution networks), In: Časopis EE , VOL XIII. Special number, 2009, 1-4

Radio Networks and Wireless Communications

chairman:

Victor RANGEL

Speech Traffic Load Sharing Between GSM and WCDMA by Using Directed Retry to GSM Feature

Hakan KAVLAK
Network Operations
Turkcell İletişim Hizmetleri A.Ş.
Ankara, Turkey
hakan.kavlak@turkcell.com.tr

H. Gökhan İLK
Electronics Engineering Department
Ankara University
Ankara, Turkey
ilk@ieee.org

***Abstract*—The most challenging problem in between GSM and WCDMA networks can be load sharing. In most cases voice is carried by GSM and data is carried by WCDMA. In Idle mode most operators want subscribers to stay in WCDMA network, so whenever the traffic load is increase, capacity utilization of WCDMA also increases while GSM network utilization still stays in low level. Directed Retry to GSM enables to share voice traffic from WCDMA to GSM. This helps equally distributed networks without losing any degradation in key performance indicators such RAB establish success, drop or user throughput. This feature is useful whenever the WCDMA user need more channel element for HSPA and GSM network utilization is sufficiently low.**

Keywords: Network load optimization, WCDMA to GSM load sharing

I. INTRODUCTION

The major challenge in today's GSM and WCDMA networks is load balancing. That is when the speech traffic in WCDMA is increased data users suffering due to low throughput. Therefore if speech can be transferred from WCDMA to GSM at a certain threshold, user throughput in WCDMA will have a better performance and return and in return GSM network won't stay idle.

Direct Retry to GSM (Global System for Mobile Communications) can be used in when there is a co-existing GSM RAN (Radio Access Network) neighbors', excess traffic in a WCDMA (Wideband Code Division Multiple Access) cell may be off-loaded to GSM, with the following benefits:

Direct Retry to GSM feature allows a WCDMA cell to handle more subscribers than it is intended for, making it possible for the WCDMA RAN to be built out at a slower pace. It allows the WCDMA RAN to focus on what it does best—to provide data services. It makes it possible for the combined system to fully utilize the existing spectrum in GSM. Speech calls with no ongoing packet connections are considered for Directed Retry during RAB (Radio Access Barrier) establishment. If the UE (User Equipment) supports GSM handover and the load of the WCDMA cell exceeds a certain threshold, the WCDMA RAN will request a blind Inter-RAT (Radio Access Technology) handover to a pre-configured GSM cell via the core network. The RAN operator has considerable control over this feature. Both the load threshold and the fraction of speech calls to be diverted to GSM are configurable.

In this study, trial area is a sub-city shown in Figure 1, Merzifon of Amasya, with two WCDMA sites, MERZI and MRYNI has been chosen for a period of week, from 26th November to 3rd December. The selection criteria that has been taken into account:

MERZI and MRYNI sites have been suffering RAB fails due to lack of UL (Uplink) Hardware called RAX. The co-existing GSM sites have no TCH/SDCCH (Traffic Channel and Semi Dedicated Control Channel) congestion and have enough room to overcome the incoming speech traffic. There are no other existing WCDMA sites around the region.

The sites in Figure 1 labeled in black represents the WCDMA sites whereas the ones in red representing the GSM ones. In the Network, the default Utrancell configuration: 1st carrier for speech/R99, 2nd carrier for HS/EUL (High Speed and Enhanced Uplink)

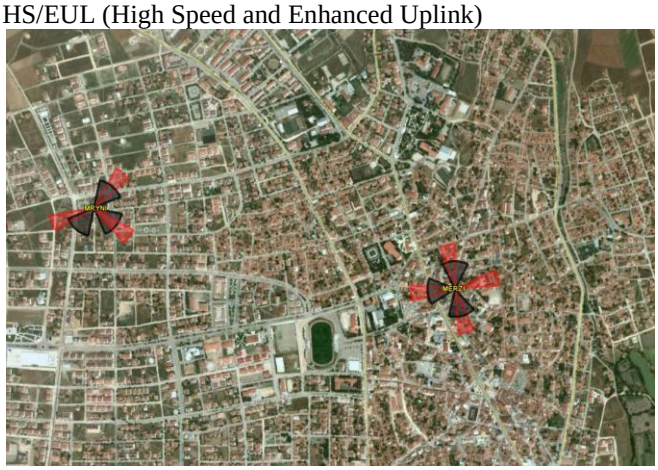


Figure 1. Trail location Merzifon with GSM sites and co-sites

Expectation of the trial is create enough capacity for WCDMA network. Before the feature implementation, it was planned to form 2 carriers identically such as; 1st and 2nd carrier for Speech/R99(Release-99)/HS/EUL with idle mode cell selection/reselection parameters equal in order not to prioritize any carrier to another. By that, with the implementation of the Directly Retry to GSM, it is possible to offload speech traffic, leave more space for data services with more resources, i.e. Channel Elements (CE).

It would be expected to have more data traffic and less RAB fails due to lack of UL Hardware. Even though many efficient coding and decoding schemes have been developed following Shannon, not until the invention of the state-of-the-art turbo codes in 1993 have we been able to approach to the channel capacity limit within just a few tenths of a dB. In other words, turbo codes have closed the significant gap between the coding gains so far achieved using the conventional coding and decoding schemes, and the channel capacity limit.

II. IMPLEMENTATION

HS-DSCH (High Speed – Downlink Shared Channel) and EUL Activation on 1st Carrier is implemented. Previously, the activation of HS-DSCH and EUL on 1st carrier was achieved in Turkcell Network.

A. Idle Mode Cell Selection/Reselection Parameters Set

Default configuration has been prioritizing 1st carrier into the 2nd previously. A subscriber attempting for HS/EUL session was selecting the 1st carrier and during RAB establishment, the session were being diverted to 2nd carrier. qOffset2sn is the Signal strength offset between source and target cells, used when the UE "cell_selection_and_reselection_quality_measure" in SIB (System Information Bits) 11/22 is set to "CPICH (Common Pilot Indicator Channel) quality called Ec/No (Carrier energy to Noise Ratio)".

B. Directed Retry to GSM Activation

Directed Retry to GSM attempts to off-load traffic from the WCDMA RAN to a co-sited GSM RAN. Speech call that has no ongoing packet connection is the only service that is targeted since it is also the only one that is safe to divert to GSM. The decision is triggered by a RAB Assignment Request from the core network. This message contains the information needed to identify the call as a speech-only call.

If a call is chosen for Directed Retry to GSM, the request for the speech RAB will be rejected with cause "Directed retry" and then a request is made to the core network to relocate the UE to a specific GSM cell, using the Inter-RAT handover procedure. This handover is a blind one since the target cell is chosen not based on UE measurements. Therefore, the target cell must be co-located with the WCDMA cell. Co-located GSM cells are assumed to have similar coverage and accessibility as their respective WCDMA cells. This algorithm is shown in Figure 2

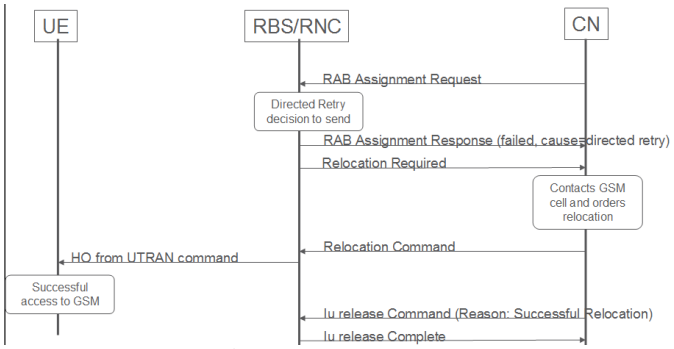


Figure 2. Flow chart of Direct Retry algorithm

loadSharingDirRetryEnabled; an RNC-wide flag for turning on the feature.

directedRetryTarget; a cell-specific parameter that specifies the Directed Retry target in terms of a cell reference to an external GSM cell.

loadSharingGsmThreshold; a cell-specific parameter that specifies the load threshold below which Directed Retry to GSM is suppressed.

loadSharingGsmFraction; A cell-specific parameter that specifies the fraction of qualified speech calls to be diverted to GSM.

The maximumTransmissionPower (MERZI11)=441 (44.1 dBm); 25.7 Watts and the pwrAdm (MERZI11)=85 %. The offloading of incoming speech calls begin when carrier transmitted power, pmTransmittedCarrierPower, of MERZI11 is above:

(maximumTransmissionPower)x (pwrAdm)x(loadSharingGsmThreshold) (1)

Thus, using eq. 1; 25.7 x 85% x 20% = 4.37 Watts. loadSharingGsmFraction =100 % means that when the above criteria is met 100% of the incoming speech RRC attempts will be directed to GSM Network. Consequently, when the RBS Tx power reaches 4.37 Watts, all of the incoming speech attempts will be directed to GSM Network. When the RBS Tx power reaches 2.19 Watts (loadSharingGsmThreshold = 10), all of the incoming speech attempts will be directed to GSM Network.

In next section main performance indicators such as accessibility to the network, call block, call drop, user throughput and the traffic.

III. PERFORMANCE OBSERVATION AND RESULT

The feature, Directed Retry to GSM, was ON from 26th November to 3rd December 2010. The performance change within this period could be seen from the following graphs. All the stats presented below spans daily average values, not busy hour values.

A. Accessibility

Here, Accessibility was observed as CS (Circuit Switch) Speech RAB Assignment Success Rate, independent of RRC Success Rate. Two Utrancell pm counters reveal the

attempts and success of speech attempts into the GSM network:
pmNoDirRetryAtt gives the total number of Directed Retry attempts and pmNoDirRetrySuccess gives the number of successful attempts. Also the directed retry success is referred as:
100x(pmNoDirRetryAtt/ pmNoDirRetrySuccess)
Usually, CS RAB assignment success rate is calculated according to eq.2

$$100*\left(\frac{pmNoRabEstablishSuccessSpeech(\ UtranCell\ )}{pmNoRabEstablishAttemptSpeech(\ UtranCell\ )}\right) \quad (2)$$

But, with the activation of the feature the formula has to be reorganized according to eq.3

$$100*\left(\frac{pmNoRabEstablishSuccessSpeech}{pmNoRabEstablishAttemptSpeech - pmNoDirRetryAtt}\right) \quad (3)$$

Here pmNoRabEstablishSuccessSpeech counts radio access barrier establishment success rate for speech and pmNoRabEstablishAttemptSpeech counts radio access barrier attempts for speech.
Since, the request for the speech RAB will be rejected with cause "Directed retry" if a call is chosen for Directed Retry to GSM. After this feature implementation, some important performance statistics are shown in below for MERZI site. It was discovered that while calculating CS Speech RAB Assignment Success Rate, the directed retry attempts, pmNoDirRetryAtt, have to be omitted as given eq. 4

$$100*\left(\frac{pmNoRabEstablishSuccessSpeech}{pmNoRabEstablishAttemptSpeech - pmNoDirRetryAtt}\right) \quad (4)$$

As the new formula representing the CS Speech RAB Assignment Success Rate of the attempts which are not directed to GSM; in other words, when RBS total output of the sectors is below 4.37 Watts for MERZI and 2.19 Watts for MRYNI, above formula presents the success of the UTRAN-accepted CS Speech RAB Attempts.
From now on, CS Speech RAB Assignment Success Rate with pmNoDirRetryAtts omitted, will be referred as CS Speech RAB Assignment Success Rate.
The total CS Speech RAB Assignment Success Rate is around 100% with 122217 CS Speech RAB Assignment Attempts, 63411 CS Speech RAB Successes with 58868 Directed Retry to GSM attempts and 55840 Directed Retry to GSM Successes. Also 48% of CS Speech RAB Attempts were directed to GSM with a success of 95%. Moreover the main factor implying the decrease for the success of Directed Retry to GSM is that MERZI11 and MERZI12 (UTRAN) does not share the same coverage region with MERZI1 (GSM) which was chosen deliberately to see the consequences. MERZI4 (GSM) has the same coverage region as MERZI11 and MERZI12. MERZI1 had 12432 Directed Retry to GSM attempts, 10428 of which is the success with a success rate of 84% shown in Figure 3 and 4.
By the way, from the graphs it can be seen that at some points the CS Speech RAB Assignment Success Rate is either negative value or more than 100%. The main reasons for that are, different ways of pegging speech RAB attempts/success from one side and Direct Retry attempts/success from another side and Soft Handover (SOHO) action that takes places between RRC connection setup and RAB establishment setup.

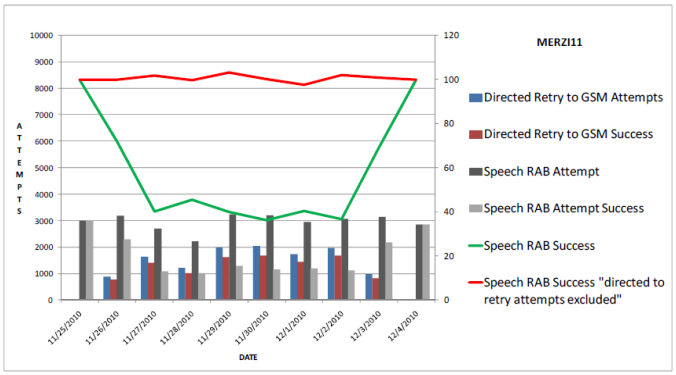


Figure 3. WCDMA Cell MERZI11 Directed Retry Success and RAB Success

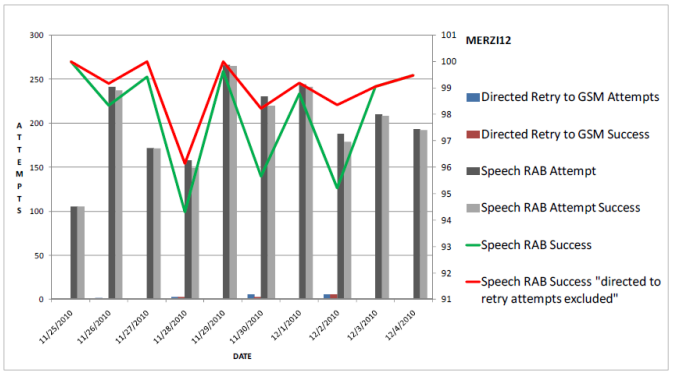


Figure 4. WCDMA Cell MERZI21 Directed Retry Success and RAB Success

B. RAB Assignment Fails Due to UL Hardware

A site suffering from RAB fails due to UL HW is said to possess non-zero peg of the counter; pmNoFailedRabEstAttemptLackUlHwBest. The best permanent solution for that is to increase the UL hardware capacity of sector with the addition of uplink channel elements and regarding license. It was aimed to gain more channel elements by offloading CS Speech Traffic into GSM network to provide more resources for data services.
As a weekly total, the MRYNI and MERZI failed 4117 times in the week of 19-25 November whereas 1363 times were experienced within the week of the feature activated, 26 November- 3 December. The greatest improvement was experienced for MERZI1 and MERZI2 from 400 fails a day to 100 fails a day. More than 1/3 improvement has been observed and shown in Figure 5. In this figure bars show the

GSM cell of MERZI1 and lines show the first and second carrier of MERZI1 WCDMA sector.

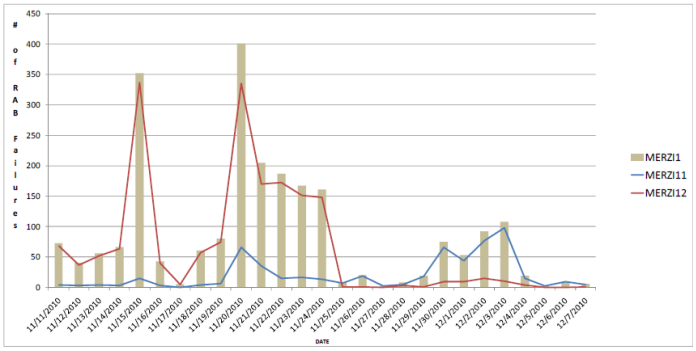


Figure 5. MERZI1 RAB Assignment Fails Due to UL Hardware

C. CS (Circuit Switch-Speech) and PS (Packet Switch-Data) Traffic

The CS Speech Traffic and HS/EUL Data Volume were investigated for about 3 weeks. The amount of the former decreased whereas the latter increased.

As expected, CS Speech Traffic of the cells with carrier1 have decreased whereas CS Speech Traffic of the cells with carrier2 have increased with the activation of HS-DSCH and EUL on first carriers and no cell selection/reselection priority between carriers. The CS Speech RAB Attempts didn't change dramatically from day to day which can be concluded as no change in user activity.

In weekly total basis, the total CS Speech Traffic decreased around 1100 Erlangs from 1650 Erlangs to 570 Erlangs. Totally 15 Erlangs of CS Speech Traffic increase in a week on the second carriers were witnessed, from 47 Erlangs to 62 Erlangs that is shown in Figure 6.

Few words can be told for HSDPA and EUL Data Volumes such that, more space for EUL and HSDPA were left, of course, with the limitations of HW, license, HS-PDSCCH codes… Moreover, the HSDPA and EUL Data Volume increased on the first carriers with the activation of those services on first carriers. Total HSDPA and EUL Data Volume increased.

As compared to the previous week the total HSDPA downloaded data for a week reached 307000 Megabytes, 32500 Megabytes greater than the previous week. In the same manner, the total EUL data uploaded for a week reached 31600 Megabytes, 2506 Megabytes greater than the week before. The experienced increase in data volume has become 4600 Megabytes for download and 360 Megabytes for upload in one day that is shown in Figure 7.

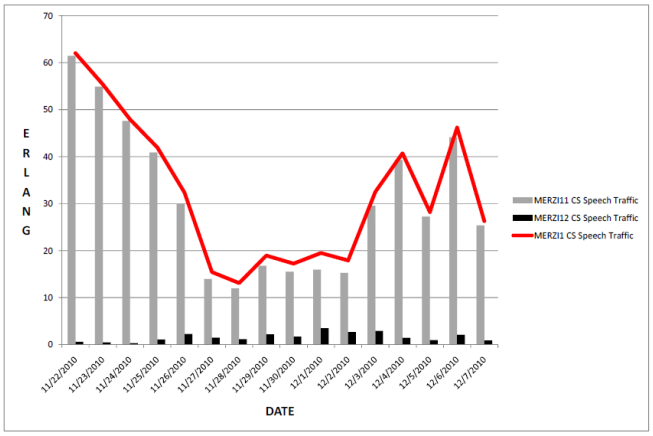


Figure 6. Speech Traffic of GSM and WCDMA Cell of MERZI1

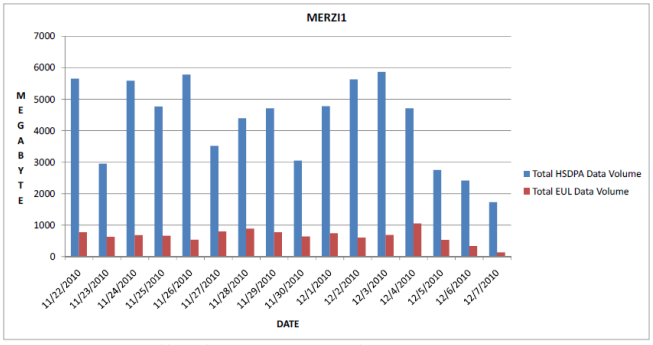


Figure 7. Data Traffic of WCDMA Cell of MERZI1

D. Retainability

There were no direct relation with the feature or HS-DSCH/EUL activation and CS/PS Retainability.

The expected result is to have similar Retainability rates before and after activation.

Just to verify the expectation, Retainability stats for CS Speech, HSDPA and EUL were presented.

Any bad performance or trend was not seen while the feature was on. As it was expected, the only tunable parameters, logics were all about accessibility but not about Retainability. The total number of drops experienced during a week was 99 with a CS Speech drop rate of 0.26% on average. The previous week, MERZI and MRYNI had 227 drops with a rate of 0.23% on average.

CS Speech Retainability changes are shown in Figure 8.

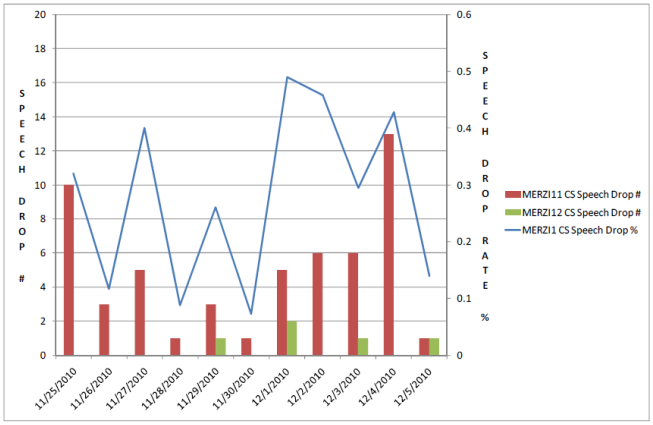


Figure 8. CS Speech Drop of GSM and WCDMA Cells of MERZI1

With the increase in HSDPA and EUL traffic, MERZI and MRYNI had more drops. Before the activation of feature and HS-DSCH/EUL on first carriers, the total numbers of HSDPA and EUL drops were 5708 and 5549 for a week, respectively. With the activation, the total numbers of HS and EUL drops for a week were experienced as 6930 and 6861 respectively.

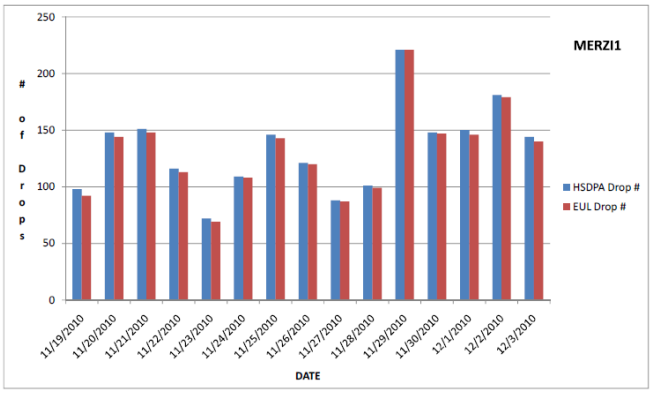


Figure 9. HSDPA and EUL Drop of WCDMA Cells of MERZI1

E. HSDPA and EUL Throughput

There is no big deal to talk about HSDPA and EUL Throughput. Although the uplink HW resources have increased in terms of more space for HSDPA and EUL traffic, both MERZI and MRYNI's busy traffic together with code, power, license limitations do not allow higher data rates.

With the activation of Directed Retry to GSM and HS-DSCH/EUL activation on first carriers, the sites MERZI and MRYNI has reached 550 kbps download and 56 kbps upload throughputs on daily average. The HS-DSCH and EUL throughputs for second carriers reached 580 kbps and 54 kbps respectively on daily average. The maximum seen throughputs became 834 kbps and 127 kbps for HS-DSCH and EUL respectively.

The HSDPA and EUL User Throughput were presented for 2 weeks is shown in Figure 10.

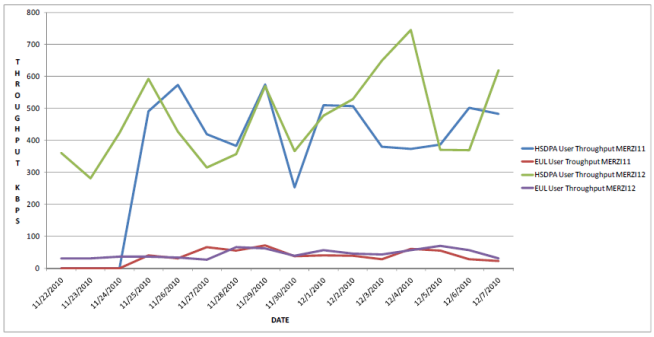


Figure 10. HSDPA and EUL Throughput of WCDMA Cells of MERZI1

When compared to the week before the activation, the change in the average number of users for HSDPA and EUL was observed. While it was 5.02, the average number of HS users became 4.25. The average number of EUL users was 5.44; previously it was 5.49 shown in Figure 11 and 12. Although the average numbers of users for both HSDPA and EUL have decreased, that does not mean that the capacity of sites also decreased. Previously mentioned KPIs, such as data volume, RAB Assignment fails imply the fact that sites had more capacity.

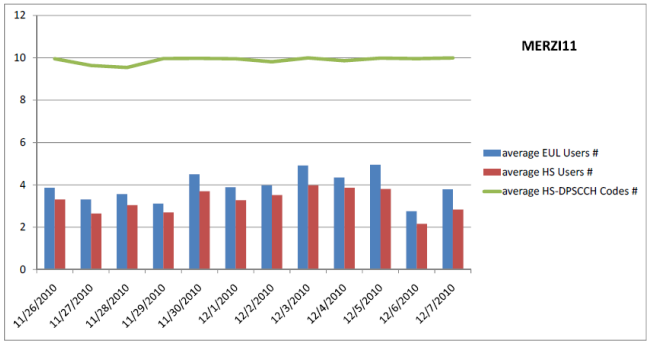


Figure 11. HSDPA and EUL Users of WCDMA Cell MERZI11

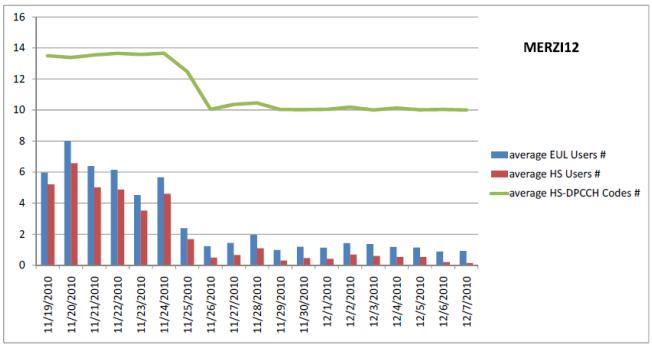


Figure 12. HSDPA and EUL Users of WCDMA Cell MERZI12

IV. CONCLUSION

The feature, Directed Retry to GSM, was tested from 26th November to 3rd December. As expected the overall performance change within this period is better to that of without Direct Retry was not used.

No performance degradation in accessibility has been observed. Moreover, failed RAB attempts due to UL HW were decreased about 35%. That is the best improvement

and aimed before the activation of the feature. Directed Retry to GSM success rate has been quite satisfying.

While CS Speech traffic were decreasing, increase in HSDPA and EUL traffic were experienced with the efficient use of downlink channelization codes and average number of users. No performance decrease in CS Speech Retainability has been seen. But the increase in HSDPA and EUL traffic lead the increase in PS Retainability.

Whole process should be considered as an experience in case of any future potential cases such as; great organizations like concert, sports match, celebration together with the successful implementation of multi carriers and R99/HSDPA/EUL load sharing features between carriers to be presented in WRAN W10 Package.

This feature is useful when high demand of HSPA High Speed Packet Access) is needed. Voice users can be carried by GSM network and all data user with better user throughput can be achieved. It is not possible to handover to GSM by using load sharing or other methods; when capacity exceeds a pre-defined threshold. However with the proposed Directed Retry to GSM.

REFERENCES

1. Ericsson Documentary, WCDMA RAN W10.1.
2. Ericsson Documentary, WCDMA Radio Access Network CPI.
3. Ericsson Documentary, LZT 123 7279 R4A WCDMA Air Interface.
4. Ericsson Documentary, WCDMA RAN Handover.
5. Laiho J., Wacker A., Novosad T., Radio Network Planning and Optimization for UMTS 2nd addition, Wiley, 2006
6. Tanner R, Woodard J. WCDMA Requirements and Practical Design, Wiley, 2004
7. Geerdes H. F, UMTS Radio Network Planning: Mastering Cell Coupling for Capacity Optimization. Mrc, 2008

Performance optimization of mobile WiMAX netwoks for VoIP streams

L. Ortiz, V. Rangel, J. Gomez.
Department of Telecommunications
UNAM, Mexico City
{lortiz, victor, javierg}@fi-b.unam.mx

R. Aquino Santos
School of Telematics
UCOL, Colima, Mexico
aquinor@ucol.mx

M. Lopez-Guerrero
Department of Electrical Engineering
UAM-I, 09340 - Mexico City
milo@xanum.uam.mx

***Abstract*—** Supporting as many VoIP (Voice over Internet Protocol) users as possible in a BWA (Broadband Wireless Access) network, using limited radio resources is a very important issue that could lead the IEEE 802.16e protocol to greater acceptance all over the world. However, VoIP service performance is affected by several characteristics defined in the BWA protocol, such as signaling overhead of control or MAP messages, ranging regions and wasted symbols, among others. In this paper we present a performance optimization to support the maximum number of VoIP calls, using codecs G.711 and G.723.1. Simulation results validate the performance optimization in terms of throughput and mean access delay for VoIP traffic.

Keywords- *IEEE 802.16e, Mobile WiMAX, VoIP, G.711, G.723.1, Performance Optimization.*

I. INTRODUCTION

The most promising solution for Broadband Wireless Access (BWA) networks is the IEEE 802.16e-2005 standard [1]. It supports mobility speeds of up to 120 km/h with an asymmetrical link structure and allows Subscriber Stations (SSs) to have a handheld form factor well suited for PDAs (Personal Digital Assistants), phones and laptops. In addition, a great deal of attention has been focused on VoIP because it is expected to be widely supported by mobile wireless networks. VoIP is a very important service because mobile users can make use of cheap voice services as compared with current mobile systems. Therefore, supporting as many VoIP users as possible in a BWA network, using limited radio resources, is a very important characteristic that could lead the IEEE 802.16e protocol to greater acceptance all over the world. However, VoIP performance is affected by several characteristics defined in the IEEE 802.16e standard, such as signaling overhead (MAP messages), ranging regions, contention, and wasted symbols due to "*rectangulation"* and "*quantization*" [1], among others.

Rectangulation is the process of allocating bandwidth resources in the downlink (DL) channel on a square or rectangular region of the frame structure. *Quantization* is the process of allocating these resources using the minimum allocation unit, denominated "*quantum map*". Considering Partial Usage of Sub-Channels (PUSC), a *quantum map* or a slot in the DL channel is one subchannel per two OFDMA (Orthogonal Frequency Division Multiple Access) symbols. A *quantum map* in the uplink (UL) channel is one subchannel per three OFDMA symbols. These settings seriously affect VoIP service performance because a VoIP packet allocation generally does not fit well in the DL and UL reservation space and usually multiple *quantum maps* are required.

Signaling overhead of MAP messages also affects VoIP services, because such overhead increases when the Base Station (BS) schedules small-sized VoIP packets. Some studies have evaluated VoIP performance taking into consideration mapping overhead [2], [3]. However, in these studies neither wasted resources in the DL channel due to *quantization* and *rectangulation* were considered nor ranging and contention signaling on the UL channel were taken into account.

A previous work presented an analytical model to evaluate VoIP performance [4]. This study takes into account the *quantization* and ranging regions in the DL and UL channels, respectively. However, wasted symbols due to *rectangulation* were not considered. The performance of some speech codecs has been evaluated in [5] and [6]. In [6] wasted symbols were considered, however they were considered as a result of variations of VoIP inter-arrival times and packet sizes. Moreover, in [5] and [6], the mapping overhead was not considered and the performance optimization of speech codecs based on OFDMA symbols for UL and DL channels was not addressed.

In this paper, we present a performance optimization of mobile WiMAX network parameters to support the maximum number of VoIP users. We make use of codecs G.711 and G.723.1 and the OPNET Modeler v.16 simulation package. The simulation model considers the issues that affect VoIP performance, such as wasted symbols due to *rectangulation* and *quantization*, signaling overhead of MAP messages, ranging and contention regions in the UL channel. In addition, a balanced amount of OFDMA symbols for each channel (UL and DL) is used in order to compensate for signaling overhead of MAP messages in the DL channel. Several modulations, codifications and encapsulations (with and without Header Suppression - HS [7]) are taken into account to evaluate the performance of G.711 and G.723.1 speech codecs under the IEEE 802.16e protocol.

The remainder of this paper is organized as follows. In Section II we present a brief description of the IEEE 802.16e protocol. In Section III and Section IV we derive theoretical and simulation models for VoIP traffic, respectively. Section V presents the performance optimization of mobile WiMAX network parameters for VoIP traffic. Finally, in Section VI we present our conclusions.

II. IEEE 802.16E PROTOCOL DESCRIPTION

The IEEE 802.16e standard uses a Medium Access Control (MAC) protocol that makes use of both Frequency Division Duplex (FDD) and Time Division Duplex (TDD). A DL channel allows transmissions from a Base Station (BS) to SSs, with PMP (Point to Multi Point) wireless access using FDD or a time signaling frame for TDD.

Multiple SSs share one slotted UL channel via TDD on demand for voice, data, and multimedia traffic. Upon receiving a bandwidth request, the BS handles bandwidth allocation by assigning UL grants based on requests from SSs. A typical signaling frame for TDD includes a DL subframe and a UL subframe. In turn, the DL subframe includes a Preamble, Frame Control Header (FCH), and a number of data bursts for SSs as depicted in Fig. 1. The Preamble is used for synchronization and equalization, and contains a predefined sequence of well-known symbols at the receiver. The FCH specifies the burst profile and length of at least one DL burst immediately following the FCH. The DL-MAP and UL-MAP are MAC management messages that include Information Elements (IEs) in order to define network access and the burst start time. These messages are broadcast by the BS following the transmission of the FCH.

Upon entering the network, each SS has to go through the Initialization and Registration setup process as defined in the IEEE 802.16e standard. Once this setup is completed, a SS can create one or more connections over which its data is transmitted to and from the BS. SSs request transmission opportunities using the UL subframe. The BS collects these requests and determines the number of OFDMA symbols (grant size) that each SS will be allowed to transmit or receive in the UL or DL subframe, respectively. This information is broadcast in the DL channel by the BS in each DL-MAP and UL-MAP. The UL-MAP message contains the IEs which describe the use of the UL subframe, such as maintenance, contention and reservation access. After receiving a UL-MAP, a SS will transmit data in the predefined reserved OFDMA symbols (UL-burst) indicated in the IEs. The DL-MAP message contains the IEs which describe the reserved zone (DL-burst) where the SS will receive its requested information.

These OFDMA symbols are transmission opportunities assigned by the BS using a Quality of Service (QoS) class, such as UGS (Unsolicited Grant Service) for CBR (Constant Bit Rate) traffic, rtPS (real-time Polling Service) for VBR (Variable Bit Rate) such as video applications, nrtPS (non real-time Polling Service) for non real-time bursty traffic, and BE (Best Effort) for Internet traffic such as email and all other non real-time traffic.

III. THEORETICAL ANALYSIS

In this section we present a theoretical analysis for the IEEE 802.16e MAC protocol when VoIP traffic (G.711 and G.723.1) is transmitted using a 20 MHz channel. The theoretical analysis can also be used to study other speech codecs (such as G.726, G.728, and G.729). We consider CBR traffic to load the network with short VoIP packets using UGS service class.

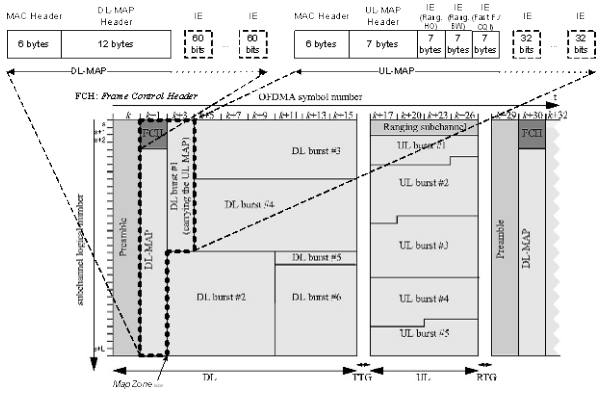


Figure 1. Frame structure for IEEE 802.16e MAC Protocol.

From Fig. 1 we can see that a DL subframe is comprised of a Preamble, a FCH, a DL-MAP message, a UL-MAP message and DL bursts. According to the standard, the Preamble and the FCH information are of constant size, but the DL-MAP and UL-MAP are of variable size.

Here the DL bursts are also constant since they are used to transport fixed-size VoIP frames. Therefore, in order to know the number of SSs supported in the DL direction ($VoIPstreams_{DL}$), we first need to compute the available number of OFDMA symbols at the PHY layer (Avl_{smbDL}) as

$$Avl_{smbDL} = \left\lfloor \frac{OFDMA_{smbDL} - Preamble_{smbDL}}{Qsmb_{DL}} \right\rfloor * Qsmb_{DL} * Data_{sbcrDL} \quad (1)$$

where $OFDMA_{smbDL}$ is the number of OFDMA symbols per subframe. $Preamble_{smbDL}$ is the number of OFDMA symbols used for the Preamble. $Qsmb_{DL}$ is the number of OFDMA symbols that comprises a *quantum map* (also termed as the minimum reservation unit) in the DL subframe. Due to the fact that data regions have to be allocated on rectangular shapes in the DL channel according to [1], we rounded down Avl_{smbDL} to multiples of *quantum maps*. $Data_{sbcrDL}$ is the number of subcarriers used for data transmission in the DL subframe. For the performance analysis we considered the default values according to [1], see Table I.

TABLE I. MAC AND PHY LAYER PARAMETERS FOR A 20 MHZ CHANNEL.

Parameter	Description	Default value	
$Frame_d$	Frame Duration.	5ms	
FFB_{smb}	Symbols for Fast Feedback/CQI	6	
FFB_{sbch}	Sub-Channels for Fast Feedback/CQI	1	
$OFDMA_{smb}$	OFDMA Symbols (see Table II)	θ[1-4]	
Rng_{smbBW}	Symbols for Ranging and BW request	1	
Rng_{smbHO}	Symbols for Ranging Handoff	2	
Rng_{sbchBW}	Sub-Channels for Ranging and BW Req	6	
Rng_{sbchHO}	Sub-Channels for Ranging Handoff	6	
		Subframe	
		UL	***DL***
$Data_{sbcr}$	Data Sub-Carriers.	1120	1440
$Qsmb$	Quantum Symbol Size.	3	2
$SbCr_{sbch}$	Sub-Carriers per Sub-Channel.	16	24

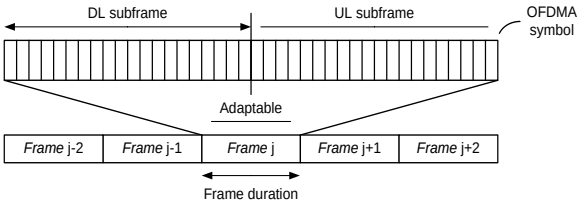


Figure 2. Variable number of OFDMA symbols for DL and UL frames.

TABLE II. OFDMA PARAMETER CONFIGURATIONS FOR UL AND DL SUBFRAMES.

Configuration	$OFDMA_{smb}$	
	UL	DL
θ0	9	38
θ1	12	35
θ2	15	32
θ3	18	29
θ4	21	26
θ5	24	23

In addition, $OFDMA_{smbDL}$ takes on values in the range of [2, 42], as defined in [1] for a 5ms frame duration ($Frame_d$). Thus, DL and UL subframes can have different configurations (θ) as shown in Fig. 2 and Table II.

Once fixed overhead has been deducted from Avl_{smbDL} (such as the *Preamble*), the number of SSs supported in the DL direction is computed as follows,

$$VoIPstreams_{DL} = \left(\frac{CIntArv_{time}}{Frame_d} \right) \left\lfloor \frac{Avl_{smbDL} - MapZone_{size} - Wstsmb_{(N)}}{SSVoIP} \right\rfloor, \quad (2)$$

here, we subtracted the variable overhead from Avl_{smbDL} ($MapZone_{size}$ and $Wstsmb_{(N)}$). $MapZone_{size}$ is the number of OFDMA symbols used by the overhead as it is shown in Fig. 1 (upper section). This amount is variable and increases proportionally to the number of accepted SSs in the network. This overhead is comprised of the fixed overhead of the Frame Control Header (FCH_{size}), the variable overhead formed by the DL MAP (Map_{sizeDL}), and the UL MAP (Map_{sizeUL}), as show in the following equation,

$$MapZone_{size} = FCH_{size} + Map_{sizeDL} + Map_{sizeUL}. \quad (3)$$

$Wstsmb_{(N)}$ in (2) is the number or wasted symbols due to *rectangulation* and *quantization*, which also increases proportionally according to the number of accepted SSs. It is computed as,

$$Wstsmb_{(N)} = \sum_{n=1}^{N} Goffset_n, \quad (4)$$

where $Goffset_n$ is the n_{th} grant offset needed by the n_{th} data allocation. $CIntArv_{time}$ is the codec inter-arrival time, which was set to 10ms and 30ms for codecs G.711 and G.723.1, respectively. *SSVoIP* is the VoIP frame size (in symbols) measured at the physical layer, considering the overhead produced by the protocol stack RTP/UDP/IP and WiMAX MAC. An example of the protocol encapsulation is shown in Fig. 3 for codec G711. Two modulations were used, QPSK1/2 (2 bits per symbol, codification=1/2) and 64-QAM3/4 (6 bits per symbol, codification=3/4). Here the constant 48 means the

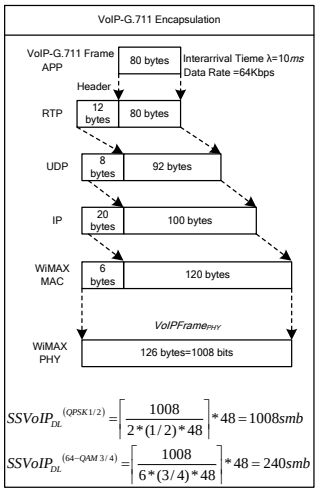


Figure 3. VoIP encapsulation for G.711 with QPSK1/2 and 64-QAM3/4.

TABLE III. NUMBER OF PHYSICAL SYMBOLS FOR *SSVoIP* WITHOUT AND WITH HEADER SUPPRESION (HS).

Codec	-HS		+HS	
	QPSK1/2	*64-QAM3/4*	*QPSK1/2*	*64-QAM3/4*
G.711	1008	240	816	192
G.723.1	528	144	336	96

quantum map size (in symbols), where *quantum map*= *Qsmb** $SbCr_{sbch}$.

Thus, *SSVoIP* can have different values depending on the modulation, codification, encapsulation type and codec used. For instance, codec G.711 can produce a VoIP frame of 80 bytes every 10 ms and codec G.723.1 can produce a VoIP frame of 20 bytes every 30 ms. This generates a data rate of 64 kpbs for G.711 and 5.3 kbps for G.723.1. Even though VoIP phones with G.711 codec can also generate a payload of 160 bytes, we used a payload of 80 bytes in order to reduce the encapsulation delay as recently studied in [8]. At the PHY layer, the reservation must be of 1008 and 528 symbols for G.711 and G.723.1 considering QPSK1/2, respectively. However, when a 64-QAM ¾ is used, the reservation at the PHY layer is reduced to 240 and 144 symbols, respectively, as show in Table III, which also includes the grant size when header suppression (HS) is used. We have considered that HS reduces the RTP, UDP and IP protocol header from 40 to 27 bytes, according to [1].

Similarly, we follow the same procedure to compute the number of SSs supported in the UL direction ($VoIPstreams_{UL}$). We computed the available number or OFDMA symbols at the PHY layer in the UL subframe (Avl_{smbUL}) taking away the ranging regions defined in [1] (as is shown in Fig. 1),

$$\begin{aligned} Avl_{smbUL} = & \left(\left\lfloor \frac{OFDMA_{smbUL}}{Qsmb_{UL}} \right\rfloor * Qsmb_{UL} * Data_{sbcrUL} \right) \\ & - Rng_{smbHO} * Rng_{sbchHO} * SbCr_{sbchUL} \\ & - Rng_{smbBW} * Rng_{sbchBW} * SbCr_{sbchUL} \\ & - FFB_{smb} * FFB_{sbch} * SbCr_{sbchUL} \end{aligned}, \quad (5)$$

where $OFDMA_{smbUL}$ is the number of OFDMA symbols in the UL subframe. We rounded down Avl_{smbUL} to multiples of *quantum maps*. $Qsmb_{UL}$ is the number of OFDMA symbols that comprises a *quantum map* in the UL subframe, and $Data_{sbcrUL}$ is the number of data subcarriers used in the UL subframe. The

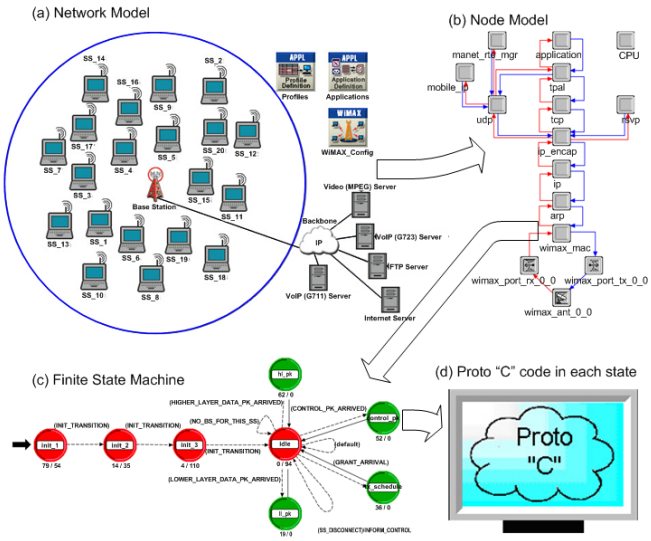


Figure 4. IEEE 802.16e simulation model.

ranging region is composed by the following parameters: Rng_{smbHO}, Rng_{sbchHO}, Rng_{smbBW}, Rng_{sbchBW}, FFB_{smb}, and FFB_{sbch}, defined in Table I. Thus, the number of supported SSs in the UL direction is computed as,

$$VoIPstreams_{UL} = \left(\frac{CIntArv_{time}}{Frame_d} \right) \left\lfloor \frac{Avl_{smbUL}}{SSVoIP} \right\rfloor . \quad (6)$$

Finally, the maximum number of VoIP streams supported in an IEEE 802.16e network is given by,

$$MaxVoIPstreams = \min(VoIPstreams_{DL}, VoIPstreams_{UL}) . \quad (7)$$

This means that we need to guarantee one VoIP stream in the UL direction and one VoIP stream in the DL direction in order to support a VoIP Call.

IV. Simulation Model

In order to validate the theoretical analysis, we implemented a WiMAX Mobile simulation model based on the OPNET MODELER package v.16. A hierarchical design was used and it is shown in Fig. 4. At the top level of the IEEE 802.16e network model are the network components (i.e., the BS and the SSs) as it is shown in Fig. 4a. The next hierarchical level, illustrated Fig. 4b, defines the functionality of an SS in terms of components such as traffic sources, TCP/UDP, IP, MAC and PHY interfaces. The operation of each component is defined by a Finite State Machine (an example of which is shown in Fig. 4c). The actions of a component at a particular state are defined in Proto-C code (see Fig. 4d).

This approach allows modifications to be applied to the operation of the IEEE 802.16e MAC protocol and different optimizations and enhancements can be tested. We simulated two scenarios, one for G.711 codec and the other for G.723.1 codec. Each scenario uses a bandwidth channel of 20 MHz, $Data_{sbcrDL}$=1440, $Data_{sbcrUL}$=1120, $Frame_d$=5ms and a repetition count of FCH and DL-MAP message equals to 4, which must be used in the worst case scenario in networks with high mobility and a high level of interference.

TABLE IV. OFDMA SYMBOL OPTIMIZATION FOR UL AND DL SUBFRAMES WITH DIFFERENT MODULATION AND CODING.

Scenery	-HS				+HS			
	QPSK1/2		*64-QAM3/4*		*QPSK1/2*		*64-QAM3/4*	
	UL	***DL***	***UL***	***DL***	***UL***	***DL***	***UL***	***DL***
G.711	18	29	18	29	21	26	15	32
G.723.1	21	26	15	32	18	29	12	35

V. Performance Evaluation

In order to achieve the maximum number of VoIP streams, we optimized the number of OFDMA symbols for the UL and DL subframes ($OFDMA_{smbUL}$ and $OFDMA_{smbDL}$). The best system configuration for the simulated scenarios is shown in Table IV.

Figures 5 and 6 show the network performance in terms of system throughput and mean access delay, respectively. In Fig. 5 we can see the throughput for the UL direction. We also found the same throughput for the DL direction, thus Fig. 5 also applies for the DL channel. In Fig. 5a, the maximum number of quality phone calls (*VoIPstreams*) supported was of 38 using G.711 codec with QPSK1/2 and without HS (-HS). This is the result of having 38 outgoing VoIP streams in the UL subframe and 38 ingoing VoIP streams in the DL subframe. In the same scenario, but considering HS (+HS), we have an increase of 45% regarding the maximum number of *VoIPstreams* which was 54. However, when 64-QAM3/4 was used, the maximum number of *VoIPstreams* increased considerably to 144 and 170 for -HS and +HS, respectively.

In Fig. 5.b we show the results for G723.1 codec, which reduces the application data rate from 64 kbps to 5.3kbps, increasing considerably the maximum number of *VoIPstreams*. With G.723.1 codec with QPSK1/2, the maximum number of *VoIPstreams* supported was of 228 without HS and 348 with HS, here the increase was of 53%. The best results were obtained with 64-QAM3/4, where we found that the maximum number of supported *VoIPstreams* could be up to 600 –HS and 738 +HS phone calls. Fig. 6 shows the mean access delay of VoIP frames in the UL direction. For G.711 codec we can that the simulated mean access delays are between 8 and 14 ms, which are under the maximum 50ms end-to-end delay (ETE) delay allowed for VoIP calls with high quality. For G.723.1 codec, the simulated mean access delays are between 18 and 26ms, which also are still below the maximum ETE delay.

VI. Conclusions

Performance optimization presented in this paper indicates that VoIP streams, under different configurations, can be supported by the IEEE 802.16e protocol. There are however, performance issues that need to be considered, as for example MOS for VoIP calls quality. However, we included this parameter and other performance issues in the extended version of this work submitted to [9].

The general trend from the results was that the system would comfortably support a number of active SSs transmitting one UL VoIP stream and one DL VoIP stream, where the maximum system throughput is obtained at the point when all available OFDMA symbols are scheduled. After that point, even a slight increase in the number of SSs results in system instability. Performance deterioration is not gradual and the packet access delay increases rapidly after this threshold if there is no admission control on the accepted traffic.

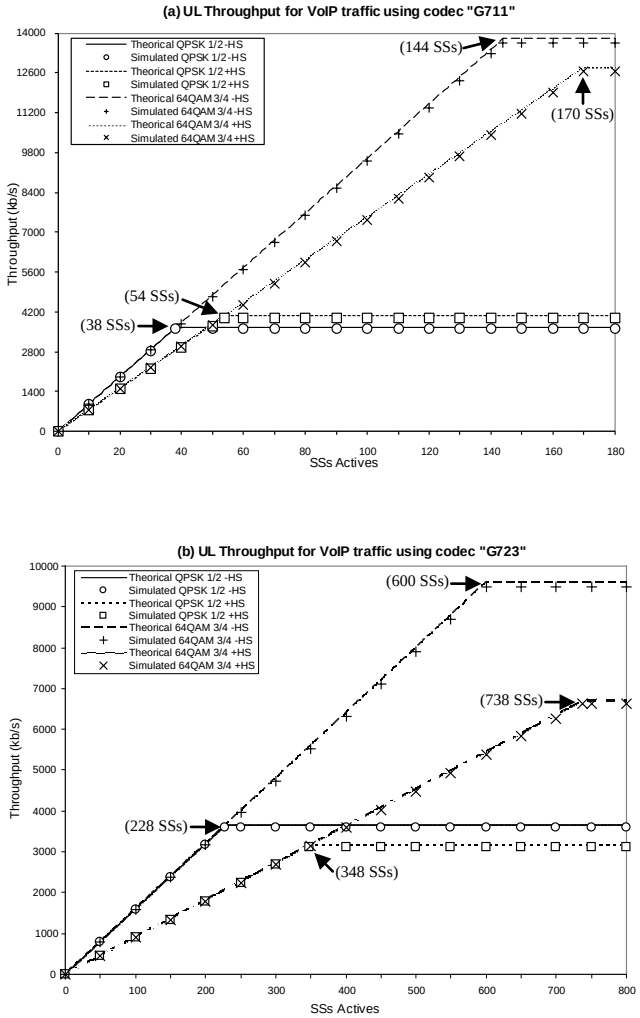


Figure 5. Maximum System Throughput of VoIP traffic in a 20 MHz Channel.

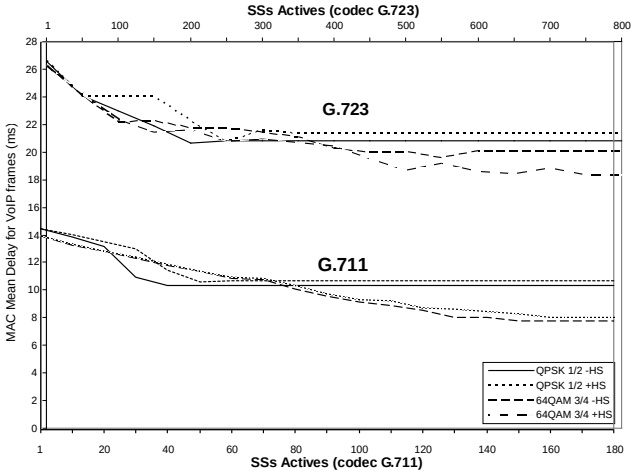


Figure 6. Mean Access Delay of VoIP traffic in a 20 MHz Channel.

ACKNOWLEDGMENT

This work was supported by DGAPA, National Autonomous University of Mexico (UNAM) under Grant PAPIIT IN108910, IN 106609, PAPIME PE103807 and CONACyT 105279, 105117.

REFERENCES

[1] IEEE 802.16e-2005, "IEEE Standard for Local and Metropolitan Area Networks - Part 16: Air Interface for Fixed BWA Systems," Amendment for PHY and MAC for Combined Fixed and Mobile Operation in Licensed Bands, Dec. 2005.

[2] Jae Woo So, "A Downlink Performance Analysis of VoIP Services Over an IEEE 802.16e OFDMA System," IEEE Communications Letters, Feb. 2007.

[3] Jaewoo So, "Performance Analysis of a Semi-Fixed Mapping Scheme for VoIP Services in Wireless OFDMA Systems," Fifth International Conference on Wireless Mobile Communications, Oct. 2009.

[4] Jae-Woo So, "Performance Analisys of VoIP Services in the IEEE 802.16e OFDMA System With Inband Signaling," IEEE Transactions on Vehicular Technology, May 2008.

[5] Kaarthick B., Yeshwenth V. J., Nagarajan N., Rajeev, "Investigating the performance of various vocoders for Fair scheduling algorithm in WiMAX," First Asian Himalayas International Conference, Nov. 2009.

[6] Howon Lee, Hyu-Dae Kim, Dong-Ho Cho, "Extended-rtPS+ Considering Characteristics of VoIP Codecs in Mobile WiMAX," IEEE 19th International Symposium on PIMRC, Dec. 2008.

[7] ETSI ES 200 800 v.1.3.1. "Digital Video Broadcasting: Interaction Channel for Cable TV Distribution Systems (CATV)," Oct. 2001.

[8] Harada N., Kamamoto Y., Moriya T., "Escaped-huffman and adaptive recursive rice coding for lossless compression of the mapped domain linear prediction residual," IEEE International Conference on Acoustics Speech and Signal Processing, Jun. 2010.

[9] Ortiz L., Rangel V., Gomez J., Aquino R., Lopez M., "Performance Evaluation of VoIP Traffic over IEEE 802.16e Protocol with Different Modulations and Codings," Submitted to Radioengineering journal, June 2011.

Wireless Sensor Network with random sending

Teresa Rajba
University of Bielsko-Biała
Department of Mathematics and Computer Science
Poland
trajba@ath.bielsko.pl

Stanisław Rajba
University of Bielsko-Biała
Department of Electrical Engineering and Automation
Poland
rajbas@ath.bielsko.pl

Abstract **The modern technology offers quite a new quality of transmitting and receiving the information due to the techniques of wireless transmission of information which are being developed. The goal of this work is to define and to test the assumptions for the accomplishment of a wireless network of sensors as a collective network (many-to-one) with random transmitting time (sending the information) of sensor-senders to WSN (Wireless Sensor Network) receiving base. We propose the accomplishment of a collective network for a certain class of applications by using PASTA system (Poisson Arrivals See Time Averages) [1], [2] for modeling WSN collective network. Unconditional probability of transmission collision occurrence has been determined. Innovativeness of this work lies in the concept of a collective radio-operated measuring network with random time of one-way transmission using one single radio frequency. The advantage of this solution is its simplicity in terms of the system and the equipment (hardware) which brings a number of further advantages related with the use of this network (a very high reliability, complete independence of information sources, easy re-configuration and mobility of the network, economy of sensor-sender power sources).**

Keywords-component; Wireless Sensor Network (WSN), Poisson Arrivals See Time Averages (PASTA system), Poisson process, probability of collision , random sending

I. INTRODUCTION

The modern solutions of wireless communication networks are very attractive. They allow the communication to be accomplished under extreme conditions. However, the accomplishment of wireless networks requires a number of conditions to be met which are much more difficult to be made than in the case of a wire network. When considering a wireless network to be produced, it is necessary to solve some problems related to it. They cover the following: the choice of network architecture (see[3],[4]) dedicated to the accomplishment of a determined task, working range of the network and autonomy of the network or its connecting to a larger communication system. The next assumption to be made is the choice of a stationary or mobile option of the network. If mobile, then its mobility must be determined somehow which is of influence on the solution of communication problems, i.e. for example, what is the migration of bases within the working range of the network. Principally, wireless networks are accomplished for two reasons, i.e. 1) network bases are assumed to be mobile, 2) we want to avoid cabling which is always a problem when building a network, with its components to be fixed. This fact is related with the problem of power supply for a base in the network. When giving up the traditional cabling there usually appears the problem of power supply for a base. Therefore, in such cases it is of great importance to solve the problem of saving the power supply source. Modern technologies of high-economy electronic systems are very helpful in this respect but the problem of power supply is also on the side of network architecture, communication procedure and spatial extensiveness. Radio transmission based communication procedure constitutes the key problem of the discussed networks.

The aim of this work is to work out and to test the communication procedure for a certain class of wireless networks. We are considering the architecture of a wireless collective network (many-to one) for measuring purposes. We assume the quantity of information sent from particular sensors not to be high. However, the number of sensors with transmitting systems, which will be referred to as „sensor-senders", may be significant. Moreover, we have only one single radio frequency (carrier wave) for the whole network. Under such circumstances there appears the problem of undisturbed transmission when each of sensor-senders must send the information to the receiving base collecting the information from particular sensors.

There is a large demand for this class of networks in different fields of science and technology. The concept of the solution presented in the paper has been motivated by the need of monitoring some climatic and biophysical parameters of a certain piece of land which was out of any technological and communication infrastructure.

Network modeling PASTA system (Poisson Arrivals See Time Averages, see[1], [2]) has been used for the need of working out the concept of operation of Wireless Sensor Network (WSN) aimed at the above mentioned purpose. Independent operation of sensor-senders is one of significant assumptions. The operation of the whole network is not dependent on a single sensor. No IT procedures are required in order to connect an additional sensor-sender. No synchronization or return channel are required for supplying the data from particular sensors to the receiving base. Network mobility should not affect its operation.

In Section 2 we have presented the assumptions for the proposed network, description of the network stochastic model and the most important features of Poisson's process which

have been used for the analysis of the proposed network. Conditional collision probability has been given in the paper [6]. In this paper the unconditional collision probability has been given (without any additional assumptions and conditions) in two cases, i.e. 1) for protocol transmission time t_p, 2) for any length of observation period. The obtained results have been plotted. Finally, we summarize results and draw conclusions.

II. NETWORK MODEL

We analyze a network consisting of n sensors which are able to send information about the measured physical magnitude on one selected radio frequency to the receiving base, quite independently of each other. Duration of communication protocol is t_p, the sensors send the information to the receiving point in randomly selected moments, every T s at an average. Beginning and cessation of transmission of a particular sensor takes place in random moments of time but these moments are relatively rare. It is a one-way transmission, i.e. from sensors to the receiving base. The sensors are completely independent of one another and their on or off state is of no influence on the operation of the network. All the sensor-senders or a part of them may be mobile provided that their senders have been left within the radio range of the receiving base. If one or more senders start sending while protocol transmission of tp time is going on from another sensor, then such a situation is called collision. Collision excludes the possibility of the correct receiving of information by the receiving base. Such a disturbed signal is ignored. The receiving base rejects the erroneous message and waits for a re-transmission to be made after the average time T. We must accept a certain loss of information in exchange for simplicity in respect of both system and equipment.

As mentioned in the Introduction, we used to modeling our wireless network a Poisson process. Poisson process is the stochastic process in which events occur continuously and independently of the one another Mathematically the process N is described by so called counter process N_t or $N(t)$ (see [5], [6]) of rate $\lambda > 0$. The counter tells the number of events that have occurred in the interval $[0, t]$ $(t \geq 0)$. N has independent increments (the number of occurrences counted in disjoint intervals are independent from each other), such that $N(t) - N(s)$ has the Poisson ($\lambda(t-s)$) distribution (mean $\lambda(t-s)$), for $t \geq s \geq 0$, $j = 0,1,2,\ldots$,

$$P\{N(t) - N(s) = j\} = e^{-\lambda(t-s)} \frac{[\lambda(t-s)]^j}{j!}. \qquad (1)$$

The rate parameter λ is the expected number of events that occur per unit time. A counting process has two corresponding random sequences, the sequence of count times (T_j) and the sequence of inter count times (U_j), such that $U_1 = T_1$ and $U_j = T_j - T_{j-1}$, for $j \geq 2$.
In the following two Propositions we give the well-known characterization of a Poisson process (see [5]).

Proposition 2.1. Let N be a counting process and let $\lambda > 0$. The following are equivalent:

a) N is a Poisson process with rate λ.
b) The inter count times $U_1, U_2, \ldots$ are mutually independent and each is exponentially distributed with parameter λ. (mean $1/\lambda$)

Proposition 2.2. Let $\lambda > 0$, $t > 0$, $n = 1,2,\ldots$. Then assuming $N_t = j$:

a) the times of n counts during $(0,t]$ are the same as n-independent uniformly distributed on the interval $[0,t]$ random variables, reordered to be non-decreasing. That is the random vector $(T_1,\ldots,T_j)$ is uniformly distributed on the set $\Omega_t = \{(t_1,\ldots,t_j) : t_1,\ldots,t_j \geq 0,\ 0 \leq t_1 \leq t_2 \leq \ldots \leq t_j \leq t\}$ with the conditional density $f(t_1,\ldots,t_j \mid N_t = j) = j!/t^j$, for $(t_1,\ldots,t_j) \in \Omega_t$, and 0 else,

b) the random vector $(U_1,\ldots,U_j)$ is uniformly distributed on the set $\Omega_t^* = \{(u_1,\ldots,u_j) : u_1,\ldots,u_j \geq 0, u_1 + \ldots + u_j \leq t\}$, with the conditional density $f(u_1,\ldots,u_j \mid N_t = j) = j!/t^j$, for $(u_1,\ldots,u_j) \in \Omega_t^*$, and 0 else.

Let us state our main assumptions. There are n identical sensors observing a dynamical system and reporting to a central location over the wireless sensor network with one radio channel. For simplicity, we assume our sensor network to be a single hop network with star topology. We also assume every node (sender-sensor, shortly sensor) always has packet ready for transmission. We assume that sensors send probe packets at Poissonian times. The average time between sending (the wake-up- times) of a sensor is T (the epoch period), and the duration of the on-time t_p (the awake interval). Assume that the wake-up- times corresponding to sensors are independent each of other . Let N be the Poisson process representing the time counter of sending sensors. Let $T_1, T_2, \ldots$ be the sending times (the wake-up- times) of sensors, $U_1, U_2, \ldots$ the inter-sending times. Then the average time between sending of sensors is T/n, the average number of sending sensors in the time interval of T length equals to n. We say that a collision occurs in the time interval of t_p length , if at least two sensors start sending within this interval. We say that a collision occurs in time interval S, if there exist at least two sensors which start

sending within this interval with the difference between the beginning of their sending time not exceeding the value of t_p.

Then the Poisson process N has the rate $\lambda = n/T$ By (1)

$$P(N_t = j) = e^{-\lambda t}\frac{[\lambda t]^j}{j!}(j = 0,1,\ldots). \quad (2)$$

Let $p(j;t_p)$ $(j = 0,1,2,\ldots)$ be the probability that the number of sensor transmissions that have occurred in the interval $[0,t_p]$ equals j. From (2), we obtain, for $j = 0,1,2,\ldots$

$$p(j;t_p) = e^{-\lambda t_p}\frac{[\lambda t_p]^j}{j!} = e^{-n\frac{t_p}{T}}[n\frac{t_p}{T}]^j / j! \quad (3)$$

In particular cases, for $j = 2,3,\ldots$, $p(j;t_p)$ can be regarded as the probability of exactly j collisions in the interval $[0,t_p]$, and consequently by the stationarity of Poisson process, on every interval $[t, t+t_p]$. $(t > 0)$.

Let A_s , $A_s^{'}$ be the events that denote the collisions occur and the lack of collisions, respectively, on the interval of s length $(s > 0)$.

Now we are going to calculate a probability of collisions on the t_p -long interval . Taking into account that a collision occurs in the time interval of t_p length, if at least two sensors start sending within this interval, by (3) we obtain:

$$P(A_{t_p}) = \sum_{j=2}^{\infty} p(j;t_p) = 1 - p(0;t_p) - p(1;t_p) =$$
$$= 1 - e^{-n\frac{t_p}{T}} - n\frac{t_p}{T}e^{-n\frac{t_p}{T}}$$

Note, that the number of sensor transmissions that have occurred in the interval $[0,t_p]$, i.e. N_t, has a Poisson distribution with the rate $\lambda t_p = n\frac{t_p}{T}$. Consequently

$$EN_t = n\frac{t_p}{T} = \sum_{j=0}^{\infty} j\,p(j;t_p) = \sum_{j=1}^{\infty} j\,p(j;t_p) \quad (4)$$

Let Y_{t_p} be the number of collisions, that have occurred in the interval $[0,t_p]$. Then we have

$$P(Y_{t_p} = 0) = p(0;t_p) + p(1;t_p) = e^{-n\frac{t_p}{T}} + e^{-n\frac{t_p}{T}}[n\frac{t_p}{T}]$$

$$P(Y_{t_p} = j) = p(j;t_p) = e^{-n\frac{t_p}{T}}\frac{(n\frac{t_p}{T})^j}{j!}, \quad j = 2,3,\ldots,$$

Consequently, by (4), we obtain:

$$E(Y_{t_p}) = \sum_{k=0}^{\infty} jP(Y_{t_k} = j) = \sum_{k=2}^{\infty} j\,p(j;t_p) =$$
$$= \sum_{k=1}^{\infty} j\,p(j;t_p) - p(1;t_p) = n\frac{t_p}{T} - e^{-n\frac{t_p}{T}}n\frac{t_p}{T} =$$
$$= (1 - e^{-n\frac{t_p}{T}})n\frac{t_p}{T}.$$

A summary of the above discussion gives the following theorem on a probability of collisions (see Fig. 1).

Theorem 2.3.

a) The probability of collisions in the interval of t_p length, is given by the formula

$$P(A_{t_p}) = 1 - e^{n\frac{t_p}{T}} - n\frac{t_p}{T}e^{-n\frac{t_p}{T}}.$$

b) The average number of collisions in the interval of t_p length is given by

$$E(Y_{t_p}) = (1 - e^{n\frac{t_p}{T}})\cdot n\frac{t_p}{T},$$

where n is the number of sensors, T is the average time between sending of a sensor and t_p is the duration time of a protocol.

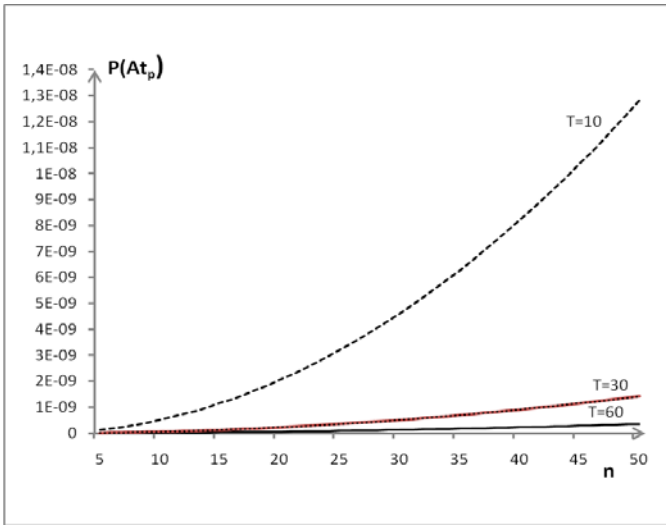


Fig. 1. Collision probability within t_p - long interval, depending on the number of sensors in the network, where $T = 10s., 30s., 60s.$

In the following theorem we give a formula for the probability of collisions $P(A_s)$, where $s > t_p$. (see Fig.2, 3, 4, 5).

Theorem 2.4. The probability of collisions in the interval of s length $(s > 0)$ is given by

$$P(A_s) = \sum_{j=2}^{\infty} e^{-n\frac{s}{T}} \frac{(n\frac{s}{T})^j}{j!}[1-(1-j\frac{t_p}{s})^j], \qquad (5)$$

where n is the number of sensors, T is the average time between sending of a sensor and t_p is the duration time of a protocol.

Proof Consider the interval $[0, s]$, where $s > t_p$. Assume $N(s) = j$, i.e. the number of sensor transmissions equals to j $(j \geq 1)$. By Proposition 2.2, the random vector $(U_1,\ldots,U_j)$ of inter transmission-times is uniformly distributed on the set

$$\Omega_t^* = \{(u_1,\ldots,u_j) : u_1,\ldots,u_j \geq 0, u_1 + \ldots + u_j \leq s\}$$

with the conditional density $f(u_1,\ldots,u_j / N(s) = j) = j!/s^j$, for $(u_1,\ldots,u_j) \in \Omega_s^*$, and 0 else. Then the conditional probability of the lack of collisions in the interval $[0, s]$, assuming $N(s) = j$ is

$$P(A_s^{'} / N(s) = j) = P(U_1 > t_p,...U_j > t_p) = (1 - j\frac{t_p}{s})^j$$

Consequently,

$$P(A_s / N(s) = j) = 1 - (1 - j\frac{t_p}{s})^j, \qquad (6)$$

for $j = 2,3,\ldots$, with the convention

$P(A_s / N(s) = 0) = P(A_s / N(s) = 1) = 0$. From (2), we have

$$P(N(s) = j) = e^{-n\frac{s}{T}} \frac{(n\frac{s}{T})^j}{j!} (j = 0,1,2,\ldots). \qquad (7)$$

By the formula of total probability we have

$$P(A_s) = \sum_{j=0}^{\infty} P(N(s) = j) \cdot P(A_s / N(s) = j).$$

Then taking into account (6) and (7), we obtain (5). This completes the proof.

Applying the Bayes theorem, which gives a mathematical representation of how the conditional probability of event A given B is related to the converse conditional probability of B given A: $P(A/B) = P(B/A) \cdot P(A/P(B)$, with $A := \{N(s) = j\}$ and $B := A_s$, we obtain the following theorem.

Corollary 2.5. The conditional probability of j sensor transmissions $(j = 2,3,\ldots).$, assuming the collisions occur, is given by

$$P(N(s) = j / A_s) = e^{-n\frac{s}{T}} \frac{(n\frac{s}{T})^j}{j!}[1-(1-j\frac{t_p}{s})^j] / P(A_s).$$

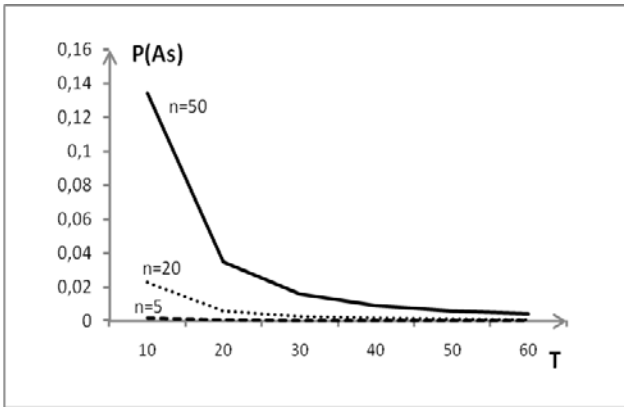


Fig.2. Collision probability depending on the average time interval between transmissions for the sensor, where $n = 5, 20, 50$. Observation time $s = 180s$ ..

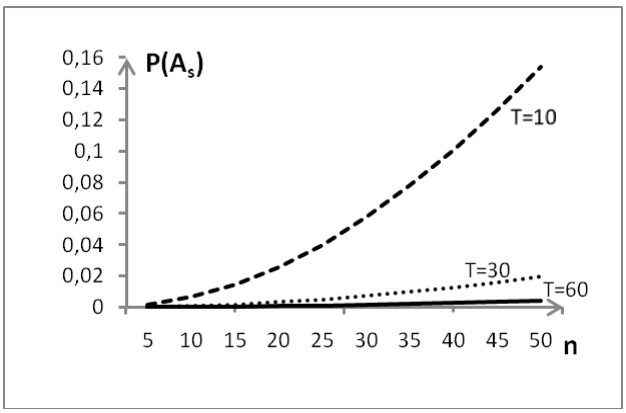


Fig.3. Collision probability depending on the number of sensors, where $T = 10s.,30s.,60s.$. Observation time s=180s..

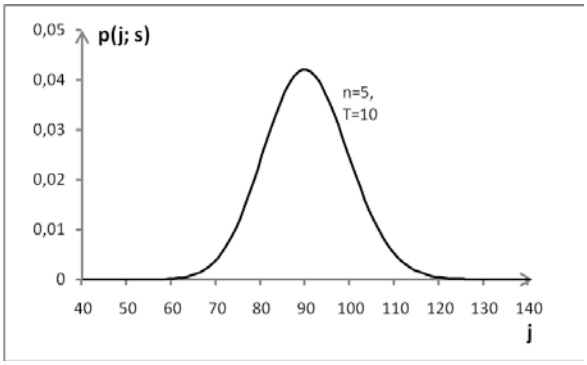


Fig. 4. Probability of j transmissions (for $n = 5, T = 10s.$) . Observation time $s = 180s$.

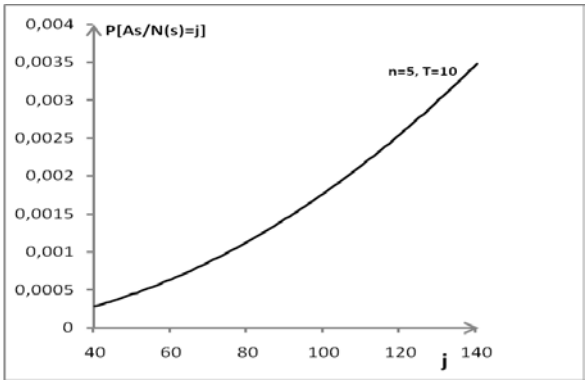


Fig. 5. Conditional probability of collision where transmission number equals to j (for $n = 5, T = 10s.$).Observation time $s = 180s$ ..

III. CONCLUSIONS

Two dependencies for the probability of a collision have been presented in the paper. The first formula determines the probability of a collision in a short protocol sending duration t_p by defining the probability of undisturbed sending of the protocol (Fig. 2). The second formula has been derived on the basis of the features of Poissson's process in respect of collision probability in any length of transmission duration. Probability of a collision has been graphically presented in diagrams as depending on the number of sensor-senders with the defined average time between subsequent message transmissions (Fig. 4); moreover, dependence has been shown on the average protocol transmission time where the number of sensor-senders is defined (Fig.3). Using the diagrams optimum values of the parameters can be found which influence the transmission correctness (n, T, t_p) . It should be determined to which extent the quality of transmission is

maintained at the present level or for which values (n, T, t_p) collision probability is growing rapidly. Range of values of collision probability can be read for optionally selected parameters: e.g. for $t_p = 3{,}2 \text{ x } 10\text{-}5$, the number of sensor-senders equal to 10 and transmitting, whereby the average transmission time every $T = 60s.$, the probability of a collision amounts to 1,65 x 10-4. It is possible to obtain so surprisingly good conditions of operation of a network with quite a large number of sensor-senders, when protocol duration t_p is by several ranges shorter than the average time interval between transmissions. Due to that the network can be qualified for a certain class of applications where such transmission parameters are acceptable. On the other hand, it can be seen that the assumptions such as one-way transmission (Simplex), maximally short transmission protocol, which does not require any synchronization or additional extension of protocol framework by additional bytes related to access procedures, control of flow, etc., enable to reach such an advantageous situation where the transmission protocol may be short. It is also extraordinarily advantageous from the point of view of battery saving which give power supply to sensor-senders (maximum saving of power supply batteries). Certainly, it is possible only where the quantity of information sent from particular sensors is small. In the analysis carried out within this paper protocol duration time has been assumed to be $t_p = 3{,}2 \text{ x } 10\text{-}5$.

On the basis of the formulas derived in the paper conditions of network correct operation can be defined for all requirements regarding the configuration of such a type of network (n, T, t_p) (collision probability). Such a surprisingly good result of transmission quality in the proposed network solution is possible only for random transmission times. There is a certain similarity to the random access control by means of CSMA method (Carrier Sense Multiple Access) in Ethernet computer network.

[1] F. Baccelli, S. Machiraju, D. Veitch, and J. Bolot: The Role of PASTA In Network Measurement, Computer Communication Review, Proceedings of ACM Sigcomm 2006 36(4):231-242, 11-15 Sept 2006

[2] F. Baccelli, S. Machiraju, D. Veitch, and J. Bolot: On Optimal Probing for Delay and Loss Measurement, in: ACM Internet Measurement Conference (IMC'07), :291-302, 24-26 Oct 2007

[3] I. Chatzigiannakis, A. Kinalis, S. E. Nikoletseas: Efficient data propagation strategies in wireless sensor networks using a single mobile sink. Computer Communications 31(5): 896-914 (2008)

[4] D. Culler, D. Estrin, and M. Srivastava. Overview of sensor networks. IEEE Computer, 37:4149, August 2004.

[5] W. Feller, Wstęp do rachunku prawdopodobieństw, PWN, Warszawa 1978

[6] T. Rajba, S. Rajba, Wireless sensor convergecast based on random operations procedure, PAK, Vol. 56 nr 3/2010, p. 255-258.

Analysis of IEEE 802.11e EDCF Media Access Mechanism

Petr Machnik
Department of Telecommunications
VSB – Technical University of Ostrava, FEECS
Ostrava, Czech Republic
petr.machnik@vsb.cz

***Abstract*—This paper describes a mathematical model of the Enhanced Distributed Coordination Function (EDCF), a novel media access mechanism used in the IEEE 802.11 technology. For this purpose, Markov Chains are employed. Based on this model, a mathematical expression of the normalized saturation throughput is derived. Furthermore, series of simulation results are shown that are focused on the dependency of the throughput and media access delay on the load for diverse EDCF parameters. Finally, the influence of the minimal and maximal contention window on the throughput and media access delay is analyzed.**

Keywords- access mechanism; EDCF; IEEE 802.11e; QoS

I. INTRODUCTION

The IEEE 802.11 technology (commonly known as WiFi) is nowadays a very popular technology that is used to create wireless local area networks or for a wireless Internet access. At the beginning, this type of networks did not distinguish various traffic classes. Later, the accent was put on the Quality of Service (QoS) support so that delay sensitive applications like voice or real-time video could be reliably transmitted by this technology.

To improve QoS, a novel standard IEEE 802.11e was published in the year 2005. One of the most important contributions of this standard is a definition of two new media access mechanisms called Enhanced Distributed Coordination Function (EDCF) and Hybrid Coordination Function (HCF). The EDCF mechanism is an improvement of an older access mechanism called Distributed Coordination Function (DCF). HCF uses both the contention-based channel access and the contention-free channel access by combining features of EDCF and PCF (Point Coordination Function). When PCF is used, an access point polls stations that can transmit data. DCF and PCF are original access mechanisms defined in the IEEE 802.11 standard. In the most practical implementations of 802.11 networks, DCF or EDCF are used because of its simplicity and efficiency.

II. EDCF MEDIA ACCESS MECHANISM

The main contribution of EDCF in comparison with DCF is the use of four access categories. The principle of EDCF is very similar to DCF with the difference that access to the shared medium is influenced by many parameters, which are distinct for individual access categories. Fig. 1 illustrates the functioning of EDCF. Within EDCF, AIFS (Arbitration Interframe Space) is used as interframe space, which is at least as large as DIFS (DCF InterFrame Space). The length of AIFS is different for different access categories. The longer is AIFS, the lower is the priority of frames.

The minimum value of the contention window CW_{min} can be also selected for each access category. CW_{min} is used to initialize the CW value. A random number between 0 and CW is chosen for each station as a back-off period otherwise two or more stations could start transmitting at the same moment after AIFS elapsing. The higher is CW_{min}, the lower is the priority of frames.

The CW_{max} value is the maximum possible value for CW on a per-category basis.

The fourth parameter dependent on the access category is the persistence factor (PF), which is used in case of an unsuccessful attempt to send data to calculate a new CW value according to the equation 1:

$$CW_{new}[AC] = ((CW_{old}[AC] + 1) \cdot PF[AC]) - 1 . \quad (1)$$

If the PF value equals 1 then the CW value does not change. If the PF value equals 2 then the new CW value is approximately twice the old CW (as well as in DCF). With increasing PF, the CW growth rate after each unsuccessful attempt to transmit data raises. The default value of PF is 2 for all access categories.

Within each station, four access categories have independent transmission queues. These queues behave as virtual stations with the described parameters determining their ability to transmit. A situation when two or more queues are trying to send their data simultaneously is called a virtual collision. The transmit opportunity (TXOP) is given to the category with the highest priority of the colliding access categories and the others back off as if a collision on the medium occurred.

More information about the EDCF mechanism can be found in [1], [2], [3].

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 218086.

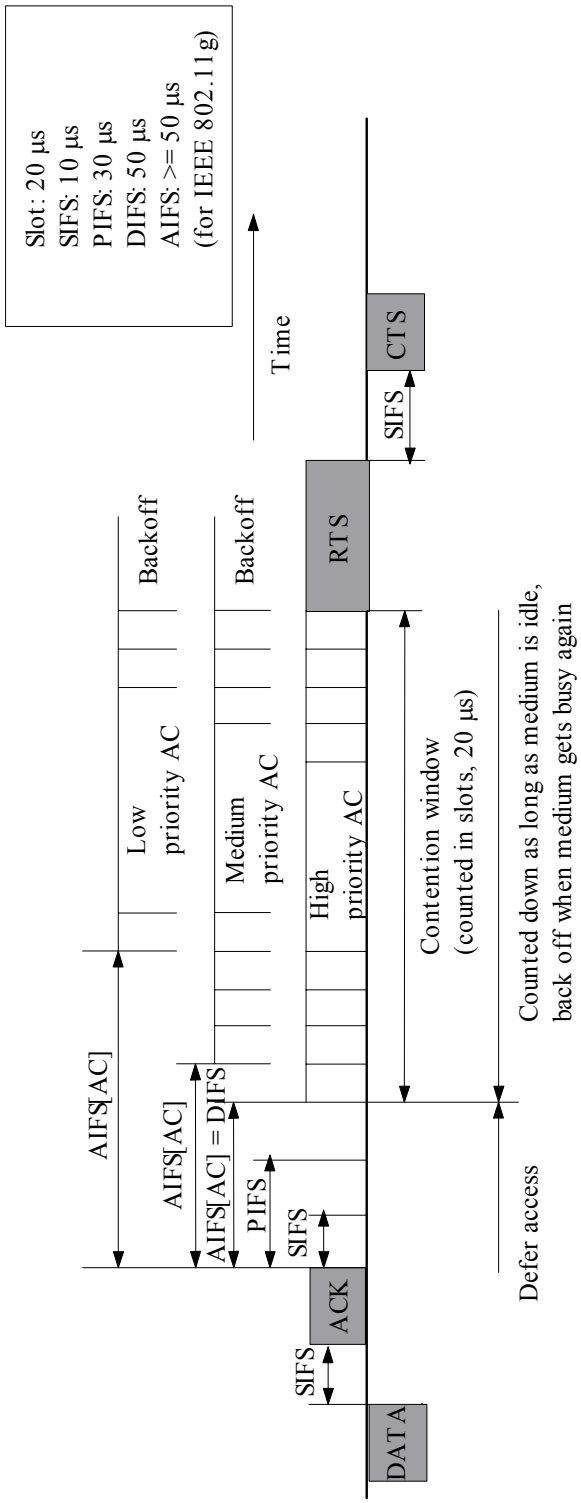


Figure 1. EDCF media access mechanism principle [2]

III. MATHEMATICAL MODEL OF EDCF

The following mathematical model of the EDCF access mechanism is based on the models of the DCF and EDCF mechanism described in [4], [5], [6] and [7].

The most suitable tool to model the EDCF access method is a two-dimensional discrete-time Markov chain (see fig. 2).

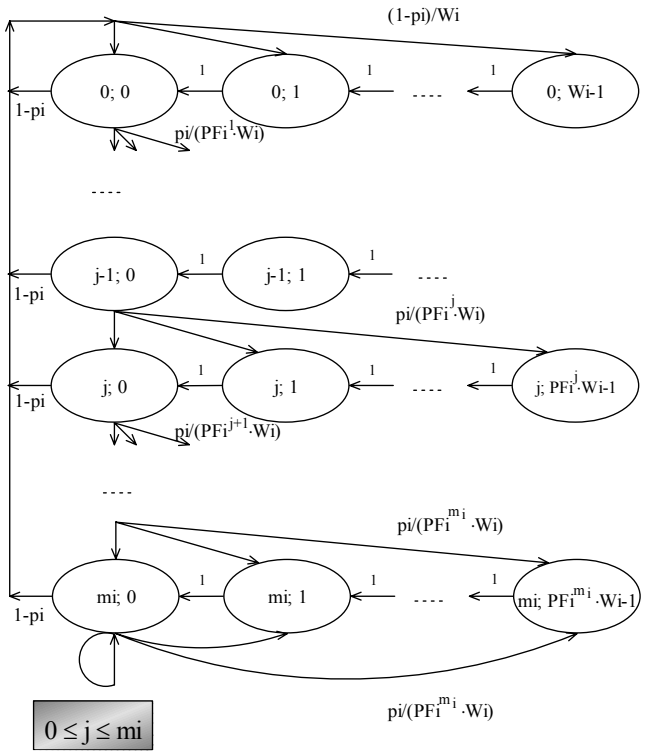


Figure 2. Scheme of an EDCF model by a two-dimensional discrete-time Markov chain

Let us assume that the channel conditions are ideal and that the system operates in saturation which means that a fixed number of stations always have a packet available for transmission.

Let N_i represents a number of traffic flows of a traffic type i ($i = 1, 2, ..., L$) belonging to an access category of a type i. L is a total number of access categories and is equal to 4 according to the IEEE 802.11e standard. Let $b_i(t)$ be the stochastic process representing the back-off time counter for a given access category of a type i that generates a pseudorandom number from the set (0, 1, …, $CW_{min,i}$-1). Moreover, let us define for convenience $W_i = CW_{\min,i}$ as the minimum contention window for an access category i. Then $CW_{\max,i} = PF_i^{m_i} \cdot W_i$ where m_i is the maximum back-off stage that leads to an increase of the contention window, i.e. until $CW \leq CW_{max,i}$. Let $s_i(t)$ be the stochastic process representing the back-off stage (0, 1, ..., m_i) for a given access category of a type i. Individual states in the two-dimensional discrete-time Markov chain are defined by couples of integers $\{s_i(t); b_i(t)\}$.

The main approximation in this model is that, at each transmission attempt for an access category of a type i, regardless of the number of retransmissions suffered, each packet collides with constant and independent probability p_i. This assumption is correct if values W_i and N_i are high.

Transition probabilities of this process are as follows:

$$
\begin{aligned}
&P\{j,k|j,k+1\}=1 \quad , \quad j\in\{0,\mathrm{K}\,,m_i\}, \quad k\in\{0,\mathrm{K}\,,PF_i^j\cdot W_i-2\},\\
&P\{0,k|j,0\}=\frac{1-p_i}{W_i} \quad , \quad j\in\{0,\mathrm{K}\,,m_i\}, \quad k\in\{0,\mathrm{K}\,,W_i-1\},\\
&P\{j,k|j-1,0\}=\frac{p_i}{PF_i^j\cdot W_i} \quad , \quad j\in\{1,\mathrm{K}\,,m_i\}, \quad k\in\{0,\mathrm{K}\,,PF_i^j\cdot W_i-1\},\\
&P\{m_i,k|m_i,0\}=\frac{p_i}{PF_i^{m_i}\cdot W_i} \quad , \quad k\in\{0,\mathrm{K}\,,PF_i^{m_i}\cdot W_i-1\},
\end{aligned}
\tag{2}
$$

where it stands

$$
P\{j_1,k_1|j_0,k_0\}=P\{s_i(t+1)=j_1,b_i(t+1)=k_1|s_i(t)=j_0,b_i(t)=k_0\}. \tag{3}
$$

Further, stationary distributions are described:

$$
\begin{aligned}
&q_{j,0}=p_i^j\cdot q_{0,0} \quad , \quad j\in\{0,\mathrm{K}\,,m_i-1\},\\
&q_{m_i,0}=\frac{p_i^{m_i}}{1-p_i}\cdot q_{0,0}, \quad \text{because } q_{m_i,0}=p_i\cdot q_{m_i,0}+p_i\cdot q_{m_i-1,0},\\
&q_{j,k}=\frac{PF_i^j\cdot W_i-k}{PF_i^j\cdot W_i}\cdot q_{j,0} \quad , \quad k\in\{0,\mathrm{K}\,,PF_i^j\cdot W_i-1\},\\
&\text{where } q_{j,k}=\lim_{t\to\infty}P\{s_i(t)=j,b_i(t)=k\}.
\end{aligned}
\tag{4}
$$

The value of $q_{0,0}$ can be derived by this way:

$$
\begin{aligned}
&1=\sum_{j=0}^{m_i}\sum_{k=0}^{PF_i^j\cdot W_i-1}q_{j,k} \quad , \quad m_i\in\{0,1,\mathrm{K}\,\},\ W_i\in\{1,2,\mathrm{K}\,\},\\
&1=\sum_{j=0}^{m_i}\left(q_{j,0}\cdot\sum_{k=0}^{PF_i^j\cdot W_i-1}\frac{PF_i^j\cdot W_i-k}{PF_i^j\cdot W_i}\right) \quad , \quad W_i\geq 1,\\
&\text{for each } j \text{ such as } 0\leq j\leq m_i, m_i\geq 0, \text{ it stands}
\end{aligned}
\tag{5}
$$

(by using an expression for a sum of arithmetic progression items)

$$
\begin{aligned}
&1=\sum_{j=0}^{m_i}q_{j,0}\cdot\frac{\left(PF_i^j\cdot W_i+1\right)}{2},\\
&1=\sum_{j=0}^{m_i-1}\left[q_{0,0}\cdot p_i^j\cdot\frac{\left(PF_i^j\cdot W_i+1\right)}{2}\right]+q_{0,0}\frac{p_i^{m_i}}{1-p_i}\cdot\frac{\left(PF_i^{m_i}\cdot W_i+1\right)}{2},\\
&\text{and because of } \sum_{j=0}^{m_i}q_{j,0}=\frac{q_{0,0}}{1-p_i},\\
&1=\frac{q_{0,0}}{2}\cdot\left[W_i\cdot\left(\sum_{j=0}^{m_i-1}(PF_i\cdot p_i)^j+\frac{(PF_i\cdot p_i)^{m_i}}{1-p_i}\right)+\frac{1}{1-p_i}\right],
\end{aligned}
\tag{6}
$$

(by using an expression for the sum of geometric progression items and L'Hospital rule)

$$
q_{0,0}=\begin{cases}\dfrac{2\cdot(1-p_i)}{W_i(1-p_i)m_i(PF_i\cdot p_i)^{m_i-1}+W_i(PF_i\cdot p_i)^{m_i}+1} & , \quad p_i=\dfrac{1}{PF_i},\\[2ex] \dfrac{2\cdot(1-p_i)}{W_i(1-p_i)\dfrac{1-(PF_i\cdot p_i)^{m_i}}{1-PF_i\cdot p_i}+W_i(PF_i\cdot p_i)^{m_i}+1} & , \quad p_i\neq\dfrac{1}{PF_i}.\end{cases} \tag{7}
$$

The probability τ_i that a back-off entity of a type i is transmitting in a generic slot time is calculated by the summation of all stationary distributions $q_{j,0}$.

$$
\begin{aligned}
&\tau_i=\sum_{j=0}^{m_i}q_{j,0}=\frac{q_{0,0}}{1-p_i}=\\
&=\frac{2\cdot(1-PF_i\cdot p_i)}{W_i(1-p_i)\left[\,1-(PF_i\cdot p_i)^{m_i}\,\right]+W_i(PF_i\cdot p_i)^{m_i}(1-PF_i\cdot p_i)+1-PF_i\cdot p_i}
\end{aligned}
\tag{8}
$$

The probability of a collision can be expressed as:

$$
p_i=1-(1-\tau_i)^{N_i-1}\prod_{h=1,h\neq i}^{L}(1-\tau_h)^{N_h}. \tag{9}
$$

Values of τ_i and p_i can be numerically calculated from two previous equations.

Following probabilities and duration times can be derived by using previous calculations:

$$
\begin{aligned}
&P_{idle}=\prod_{i=1}^{L}(1-\tau_i)^{N_i},\\
&P_{busy}=1-P_{idle}=1-\prod_{i=1}^{L}(1-\tau_i)^{N_i},\\
&P_{success,i}=\begin{cases}0, & P_{busy}=0,\\ \dfrac{1}{P_{busy}}\cdot N_i\cdot\tau_i\cdot(1-\tau_i)^{N_i-1}\cdot\prod_{h=1,h\neq i}^{L}(1-\tau_h)^{N_h}, & P_{busy}\neq 0,\end{cases}\\
&P_{collision,i}=1-P_{success,i},
\end{aligned}
\tag{10}
$$

where P_{idle} is probability that the transmission medium is idle,

P_{busy} is probability that the transmission medium is busy,

$P_{success,i}$ is probability that the transmission attempt of a category i is successful,

$P_{collision,i}$ is probability that the transmission attempt of a category i is unsuccessful.

T_{idle} - *slot time* (It depends on the type of the physical layer.)

$$T_{success,i} = \begin{cases} T_{PLCP} + \delta + T_{SIFS} + T_{ACK} + \delta + T_{AIFS,i} \\ \text{(without RTS/CTS)}, \\ (T_{RTS} + \delta + T_{SIFS} + T_{CTS} + \delta + T_{SIFS}) + T_{PLCP} + \\ + \delta + T_{SIFS} + T_{ACK} + \delta + T_{AIFS,i} \quad \text{(with RTS/CTS)}. \end{cases} \tag{11}$$

$$T_{collision,i} = \begin{cases} T_{PLCP} + \delta + T_{AIFS,i} & \text{(without RTS/CTS)}, \\ T_{RTS} + \delta + T_{AIFS,i} & \text{(with RTS/CTS)}. \end{cases}$$

where T_{idle} is a period when the transmission medium is idle,

$T_{success,i}$ is a period of an successful transmission of a category i,

$T_{collision,i}$ is a period of an unsuccessful transmission of a category i,

T_{PLCP}, T_{RTS}, T_{CTS} represent periods that are necessary to send a PLCP, RTS, CTS frame respectively,

T_{AIFS}, T_{SIFS} represent corresponding interframe spaces,

T_{DATA} represents a period used for transmitting MAC sublayer payload,

δ is a propagation delay.

Finally, we can express normalized saturation throughput for the EDCF media access mechanism:

$$S = \sum_{i=1}^{L} S_i = \frac{P_{busy} \cdot \sum_{i=1}^{L} P_{success,i} \cdot T_{data}}{P_{busy} \cdot \sum_{i=1}^{L} \left(P_{success,i} \cdot T_{success,i}\right) + P_{busy} \cdot \sum_{i=1}^{L} \left(P_{collision,i} \cdot T_{collision,i}\right) + P_{idle} \cdot T_{idle}}. \tag{12}$$

$S \in (0;1)$.

The saturation is normalized with regard to the physical layer technology. The higher is the saturation throughput the more effective is time spent for transmitting of useful data compared with time spent because of an overhead, a collision recovery and a back-off delay. Probabilities in the equation serve as influence weights of time items on the throughput.

IV. Simulation Results

All simulations were made by the OPNET Modeler network simulation program. The simulated network topology consisted of 20 stations and one access point (see fig. 3). The distance of stations from the access point was about 20 meters without any obstacles. The IEEE 802.11g technology was used at the physical layer. All devices supported the EDCF mechanism and the traffic differentiation. Wireless stations were divided equally for five among four groups. Each group generated traffic belonging to one of access categories. The length of packets created by stations was 1000 bytes. Packets were generated by stations with the constant probability. Their interarrival time was changed to influence the offered load. Three sets of EDCF parameters for individual access categories were used to demonstrate their influence on the data throughput and media access delay. Their values are shown in tables 1 – 3. The RTS/CTS mechanism was not used in these simulations.

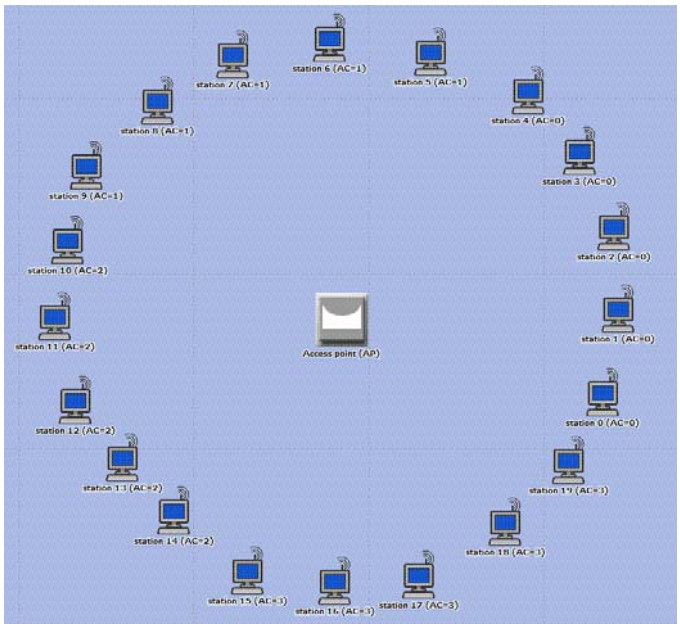


Figure 3. Network topology for simulations

TABLE I. EDCF Parameters Set No. 1
(recommended and usually default values)

Access Category	EDCF Parameters			
	CW_{min}	CW_{max}	AIFS[μs]	TXOP[ms]
0	15	1023	150	0
1	15	1023	70	0
2	7	15	50	3
3	3	7	50	1.5

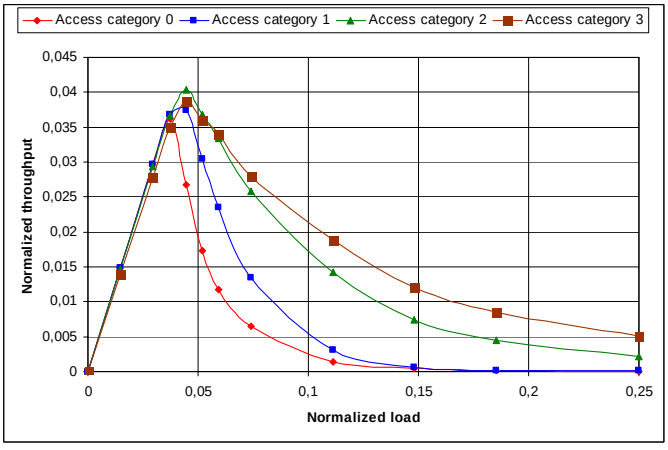


Figure 4. Graph for comparison of EDCF access categories (set no. 1)

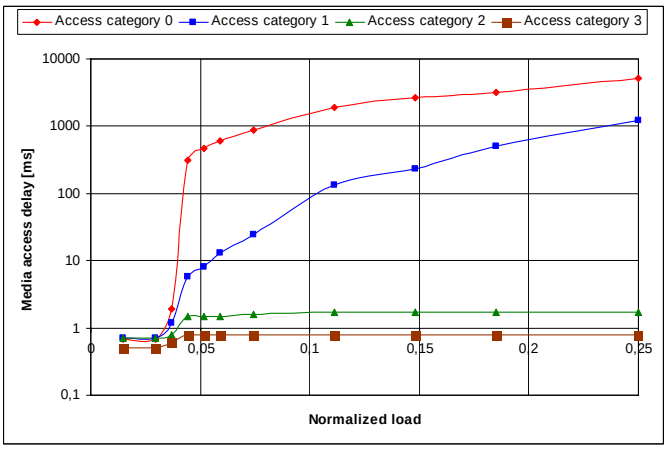


Figure 5. Graph for comparison of EDCF access categories (set no. 1)

Results of the first set of simulations arise from the graphs in fig. 4 and 5. The first graph describes the dependency of the throughput on the load for individual access categories. A peak of the throughput is possible to see herein. Further growth of the load induces more collisions and thereby lower throughput. The second graph shows the dependency of the media access delay on the load for individual access categories.

In the second set of simulations, EDCF parameters with very close values were used (see table 2). As a result, throughput values for individual access categories are nearly the same (see fig. 6). The media access delay for low priority access categories is lower compared with the previous simulations (see fig. 7).

TABLE II. EDCF PARAMETERS SET NO. 2

Access Category	EDCF Parameters			
	CW_{min}	CW_{max}	*AIFS[μs]*	*TXOP[ms]*
0	15	127	50	0
1	7	127	50	0
2	3	15	50	3
3	3	7	50	1.5

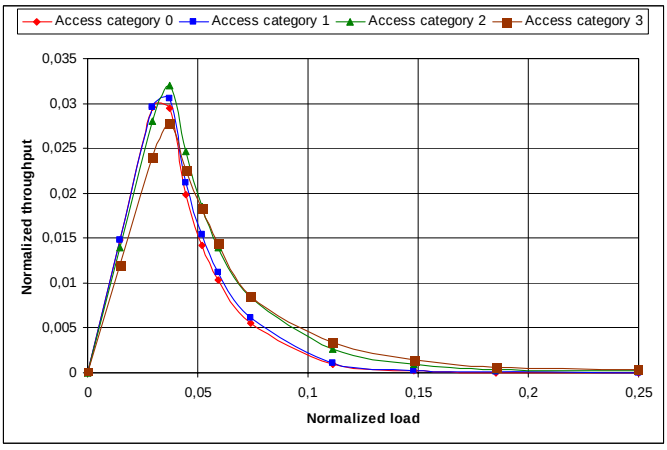


Figure 6. Graph for comparison of EDCF access categories (set no. 2)

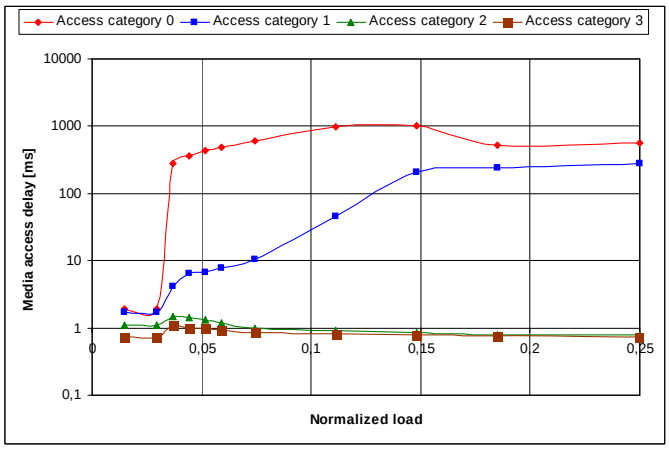


Figure 7. Graph for comparison of EDCF access categories (set no. 2)

The third set of simulations demonstrates the effect of very dissimilar values of EDCF parameters (see table 3). As can be seen in fig. 8, only access categories 2 and 3 can be effectively used to transport data during the high throughput. Within access categories 0 and 1, the data throughput is very low. The media access delay (see fig. 9) is quite high within all access categories compared with results of other simulations because of high EDCF parameters within categories 0 and 1 and many collisions of frames within categories 2 and 3 that have close values of EDCF parameters.

TABLE III. EDCF PARAMETERS SET NO. 3

Access Category	EDCF Parameters			
	CW_{min}	CW_{max}	*AIFS[μs]*	*TXOP[ms]*
0	127	4095	210	0
1	31	1023	90	0
2	3	15	50	3
3	1	7	50	1.5

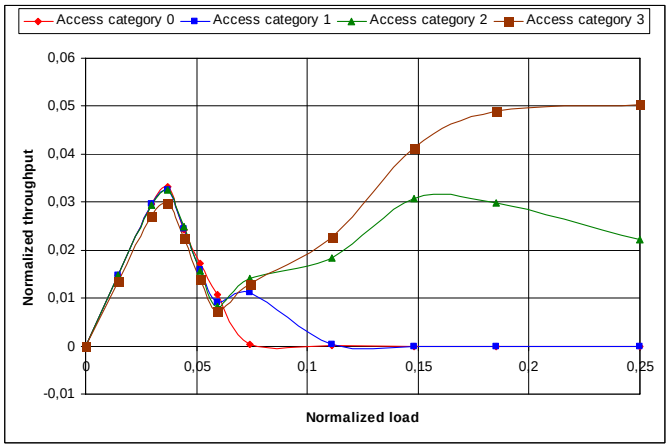


Figure 8. Graph for comparison of EDCF access categories (set no. 3)

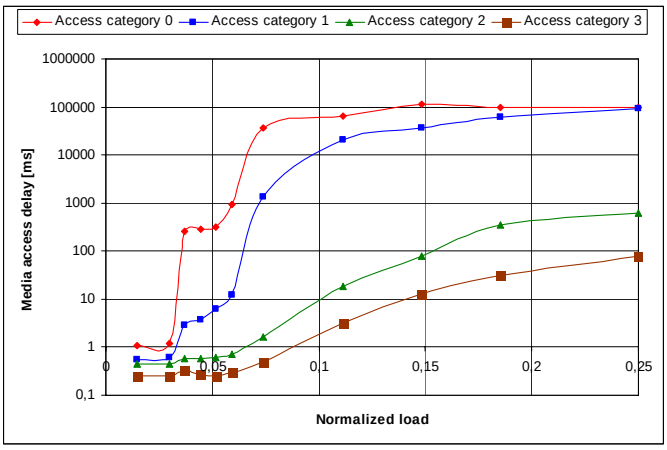


Figure 9. Graph for comparison of EDCF access categories (set no. 3)

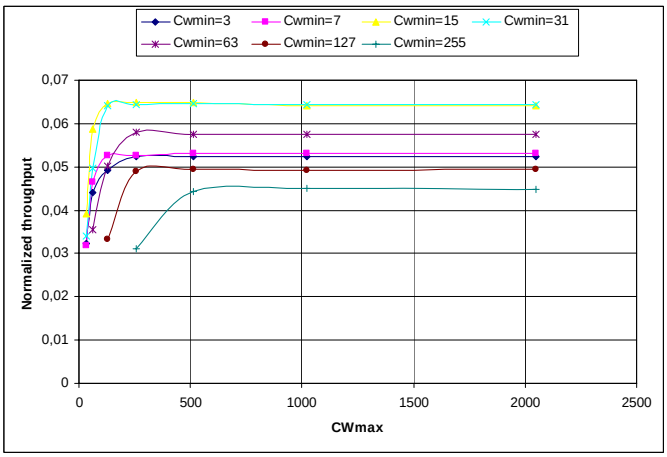


Figure 10. Graph that illustrates influence of CW_{min} and CW_{max} on throughput

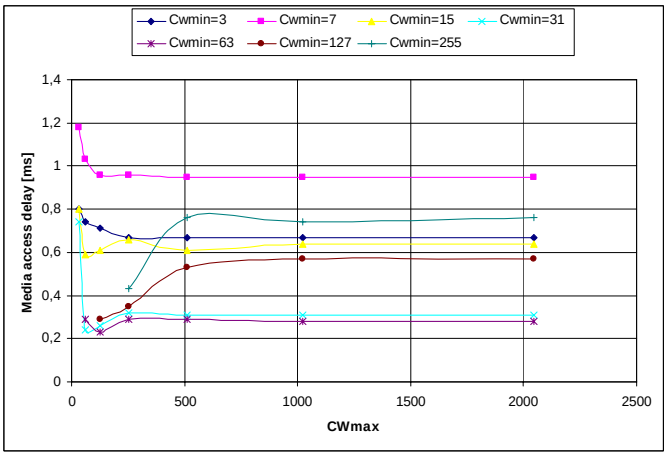


Figure 11. Graph that illustrates the influence of CW_{min} and CW_{max} on media access delay

In fig. 10 and 11, graphs are shown that illustrate the influence of the minimal and maximal contention window on the normalized throughput and media access delay. It is obvious that a strong influence of CW_{min} parameter but a small influence of CW_{max} parameter exist. However, if both parameters have close values then the throughput is low and the delay is high. Best results from all simulations (high throughput and low media access delay) were accomplished when CW_{min} = 31 and CW_{max} =127 (or higher). Simulations were performed with a traffic load 12.8 Mbit/s.

V. Conclusion

This paper described a mathematical model of the EDCF media access mechanism based on Markov chains. An equation for the normalized saturation throughput was derived thereby the influence of EDCF parameters was illustrated. Realized simulations demonstrated the dependency of the throughput and media access delay on the load for distinct access categories and for three sets of EDCF parameters. A significant difference among individual access categories was found out. The priority of traffic belonging to a specific access category is strongly affected by the used EDCF parameters values. Finally, a strong influence of the CW_{min} parameter on the throughput and media access delay was shown.

References

[1] *IEEE 802.11e Standard : Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specification - Medium Access Control Quality of Service Enhancements* [online]. 2005. Available: http://standards.ieee.org/getieee802/download/802.11e-2005.pdf.

[2] F. Ohrtman. *Voice over 802.11*. Boston: Artech House, 2004, 258 p. ISBN 1580536778.

[3] P. Roshan, J. Leary. *Wireless Local Area Network Fundamentals*. Cisco Press, 2003, 300 p. ISBN 1587050773.

[4] P. Machnik. "Mathematical Modeling and Simulation of EDCF Media Access Mechanism in IEEE 802.11 Technology." In *Proceedings 31th International Conference Telecommunications and Signal Processing 2008*. Budapest: Asszisztencia Szervező Kft., 2008. ISBN 978-963-06-5487-6.

[5] G. Bianchi. "Performance Analysis of the IEEE 802.11 Distributed Coordination Function." In *Selected Areas in Communications, IEEE Journal*. 2000, pp. 535-547. Available: http://www.sop.inria.fr/maestro/personnel/Sara.Alouf/Stage/bianchi2000.pdf. ISSN 0733-8716.

[6] B. Li, R. Battiti. "Performance Analysis of An Enhanced IEEE 802.11 Distributed Coordination Function Supporting Service Differentiation." In *Quality for All : 4th COST 263 International Workshop on Quality of Future Internet Services, QoFIS 2003*. Berlin: Springer, 2003, pp. 152-161. ISBN 9783540201922.

[7] S. Mangold. "Analysis of IEEE 802.11e and Application of Game Models for Support of Quality-of-Service in Coexisting Wireless Networks." Mainz : Wissenschaftsverlag Mainz, 2003. 269 p. ComNets Aachen University. Ph.D. Dissertation. Available: http://www.comnets.rwth-aachen.de/typo3conf/ext/cn_download/pi1/passdownload.php?downloaddata=19|/var/www/cnhome/downloads/publications/. ISBN 3860739850.

Visualisation of Best Servers Areas in GSM Networks

Roman Sebesta, Marek Dvorsky, Lukas Kapicak,
and Libor Michalek
Department of Telecommunications
VSB-Technical University of Ostrava, FEECS
Ostrava, Czech Republic
{roman.sebesta, marek.dvorsky, lukas.kapicak,
libor.michalek}@vsb.cz

Peter Scherer,
and Jan Martinovic
Department of Computer Science
VSB-Technical University of Ostrava, FEECS
Ostrava, Czech Republic
{peter.scherer, jan.martinovic}@vsb.cz

***Abstract*—The article deals with improving of basic localization techniques used in mobile networks for determining the location of mobile station. Proposed method uses network parameters which are obtained by continuous measurement in the mobile network. These parameters create a database of values necessary for the identification of areas which are covered by the base stations signal.**

***Index Terms*—Localization, GSM, Best Server, Timing Advance, Cell ID**

I. Introduction

Nowadays, there are well known various methods for determining the location position of Mobile Stations (MS) [1]. The best method for the independent MS localization (without necessary additional network modifications) is to determine MS position from the side of GSM network (NBP - Network Based Positioning).

The basic positioning method is based on the Cell Identification (Cell ID) [2] [15] [16]. An explored area with a surveyed MS is defined as a circular area around the active Base Transceiver Station (BTS) (see Fig. 1. - hatched area 1). An accuracy improvement of this method is achieved if BTS uses sector antennas. This method faces to a circular sector by reducing the circular area 1 (see Fig. 1. - squared area 2).

Father accuracy achievement of MS localization for BTS which covers areas over a distance of 550 m, can be reached by limiting area 2 on Figure 1 by network parameter Timing Advance (TA) [1] [11] shown in Fig. 1. - dotted area 3. The accuracy of this method depends on the size of the covered area, ranging from hundreds of meters to several kilometers [8]. The main quality of this method is the independence from the type of MS and a low cost implementation into the real mobile network [3] [14].

Areas covered by particular sectors of the BTS can be determined using sophisticated software tools, which are used by mobile operators, to optimize their mobile network [4]. Graphical visualization of signal coverage is based on the known signal propagation models involving the terrain profile, as well as buildings, in the surveyed area. Although those sophisticated softwares are considered as the professional tools, their outputs do not every time reflect the actual situation existing in the surveyed area. This can be detected only by measurements performed in specific reception areas. Mobile operators perform regular measurements of several network parameters with the main purpose of the network optimization. Regarding those measurements the operator disposes an extensive measurement's database that is regularly updated.

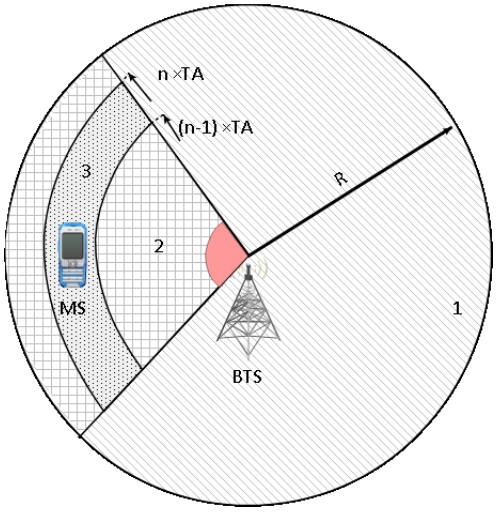


Fig. 1. Basic localization method in GSM network.

This paper describes an improvement in the basic positioning methods used in GSM cellular networks that are based on the identification of Cell ID [7] [10]. It uses measured values of network parameters [17] that contain information about the current network settings and reception conditions in the service and the six surrounding BTS [13]. Recorded information is sent by each mobile station during communication over the GSM cellular network. These parameters are consequently used as input values of the proposed algorithm for determining the Best Server area of the BTS sector.

II. The Reference Database Preparation

A. Description of a measurement device for data logging

Measurement device (NAM Logger) is used for obtaining selected parameters in the GSM network. NAM logger is

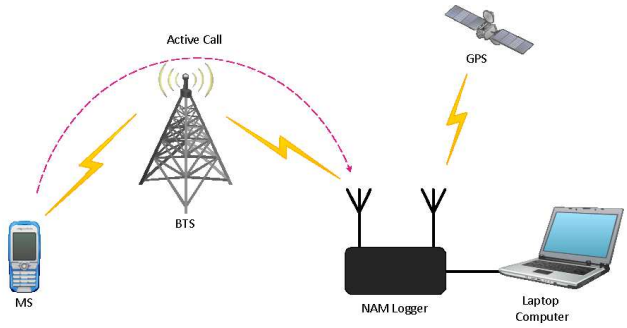


Fig. 2. Measurement tools for data logging.

composed of the following parts:

- microprocessor,
- internal flash memory,
- GSM / GPRS / EDGE modem Sony Ericsson GR47,
- GPS modul.

NAM Logger uses external GPS and GSM antennas. As the end terminal representing MS is used a common mobile phone. The active call is performed from the site of MS to NAM Logger (see Fig. 2.).

B. Description of data logging process

The actual measurement is performed in Dedicated (speech) Mode (DM). DM mode is used with reason of comparison with existing localization methods. The comparison is based on measurement a circular area with a radius corresponding to the size n x 550 m, where n is the value of TA (see Fig. 1., area 3) [5] [6]. Measurement is performed in the two following steps:

1) MS makes a Call Setup between MS and NAM Logger (call initiates MS, NAM Logger automatically answers).
2) Measured parameters are automatically saved into the internal memory of NAM Logger within a specified time period.

The data for described experiments were measured in Ostrava - Poruba and adjacent urban neighborhoods. The monitored area is approximately 10 square kilometers.

C. Description of monitored network parameters

For network parameters is used a custom mode of GSM modem called Ericsson M2M Engineering Monitoring Mode [9] [12]. Using AT command AT*EEMM modem enables read the following network parameters: Serving Cell, MCC, MNC, LAC, CellID, BSIC, Ch[RxL, C1, C2] [RxLFull, RxLSub, RxQFull, RxQSub, TA, TN]. The following parameters are recorded into the internal memory of NAM logger in pre-defined intervals:

- Current GPS position (Longitude, Latitude) for viewing the actual position where the measurements was performed.
- Cell Identification.
- Current receive level (RXLEV) measured towards the traffic channel (TCH).

Fig. 3. Application LoMo Map Viewer for visualization of the measured network parameters.

- Parameter TA for comparison of the proposed method with the original TA localization method.

Recording period is set to one sample per five seconds. The total length of measured route in the analyzed area is approximately 30 kilometers. During the monitoring the GSM and GPS antennas are placed on the roof of a car. There were made two referential databases. First data-set consists of 772 samples; second set consists of 498 samples.

D. Description of visualization by developed application

The application (LoMo Map Viewer) was developed to generally visualize all parameters of the measured network. It allows visualizing of individual measured values together with their GPS position (see Fig. 3.). This visualization allows further analysis of acquired network parameters. LoMo Map Viewer consists of several layers. Over the base map layer is applied a layer consisting of individual measurements and other objects. There is also an ability to save and load the internal database as well as an import of several file formats.

III. Method for Best Server Visualization

Area of the interest is location with measured data defined by the GPS coordinates (top left point and down right point). Set C of all Cell IDs is defined as $c_1, c_2, \ldots, c_m \in C$, where m is amout of different Cell ID in area of the interest (G) and elements of the set C are represented by name of the Cell ID. Grid is defined as squares r_{ij} from the area of the interests, $r_{ij} \in G$. A map grid is a set of lines running horizontal and vertical forming small square on a map. One square from map give a location for a specific area. Set V of inputs in one square is defined as $v_1, v_2, \ldots, v_m \in V$, where v_i is pair (c, RXLEV), where $c \in C$. For each square r_{ij} is

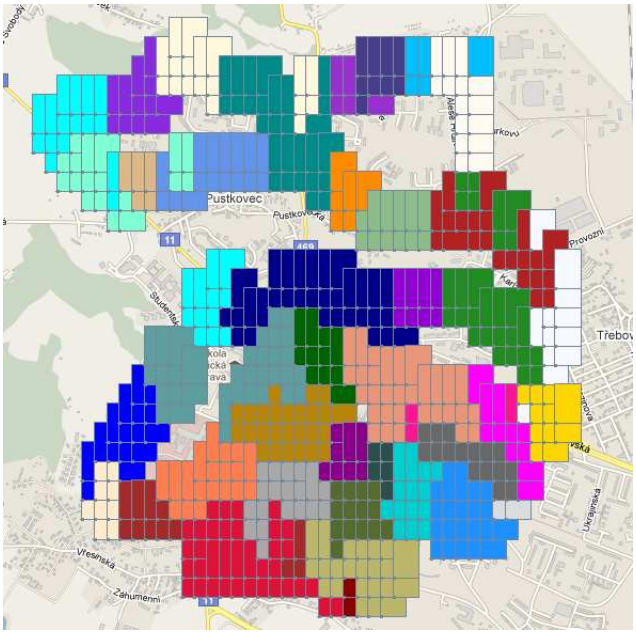


Fig. 4. Final calculated map of Best Server in LoMo Map Viewer.

Algorithm III.1 Algorithm DrawMap – function $CreateGrid$

```
function CreateGrid(lng_start, lat_start)
    lng_step ← 0.002
    lat_step ← 0.002
    for i = grid_lng_start to grid_lng_end by 1 do
        for j = grid_lat_start to grid_lat_end by 1 do
            p00 ← create 0,0 marker border point at
            lat : lat_start + (j/1000.0)
            lng : lng_start + (i/1000.0)
            p01 ← create 0,1 marker border point at
            lat : lat_start + (j/1000.0)
            lng : lng_start + (i/1000.0) + lng_step)
            p10 ← create 1,0 marker border point at
            lat : lat_start + (j/1000.0 + lat_step)
            lng : lng_start + (i/1000.0)
            p11 ← create 1,1 marker border point at
            lat : lat_start + (j/1000.0 + lat_step)
            lng : lng_start + (i/1000.0) + lng_step)
            marker ← create marker from p00, p01,
            p10 and p11
            map.Add(marker)
        end for
    end for
end function
```

defined vector $g_{ij} = (w_{ij1}, w_{ij2}, w_{ijn})$, where $n = |C|$ and the weight w_{ijk} is strongest RXLEV attribute of $v \in V$, where $v = (c_k, RXLEV)$ in square r_{ij}. Square is colored by color of best weighted Cell ID in vector g_{ij}.

Generally, the denser grid (lng_step, lat_step), the more accurate final calculation is made. With the denser grid is necessary to have a greater variance of the GPS positions measured values to achieve optimal results in visualization.

The second step is to assign these values to the appropriate X, Y positions in the coordinate's grid (see Algorithm III.2, III.3). When the values are assigned the internal database of CellID is considered. In the case of a new unique value, it is stored into the database and indicated by different color.

Algorithm III.2 Algorithm DrawMap – function $CheckPoint$

```
function CheckPoint(point p, point mp1, point mp2)
    if p.Lat <= mp1.Lat ∧ p.Lat >= mp2.Lat then
        if p.Lng >= mp1.Lng∧p.Lng <= mp2.Lng then
            return True
        else
            return False
        end if
    end if
end function
```

Algorithm III.3 Algorithm DrawMap – function $SetValues$

```
function SetValues(array points)
    for i = 0 to point_count by 1 do
        p ← points[i]
        p_id ← p.CellId
        for j = 0 to map.MarkersCount by 1 do
            mp1 ← map.Markers[j].Border[0]
            mp2 ← map.Markers[j].Border[3]
            if CheckPoint(p, mp1, mp2) then
                marker ← map.Markers[j]
                marker.Database.Add(p)
                if p_id ∉ map.ColorDatabase then
                    map.ColorDatabase.Add(p_id)
                    map.ColorDatabase[p_id] ← set color
                    p.Color ← map.ColorDatabase[p_id]
                end if
            end if
        end for
    end for
end function
```

When all points are assigned, an internal analysis explores which point of the X, Y position in the grid has the strongest RXLEV. After location the major point is the section colored by value that was assigned in the second step. This process creates a complex Best Server map (see Algorithm III.4).

The final Best Server map is visualized in LoMo Map Viewer (see Fig. 4.).

Moreover, the LoMo Map Viewer is able to visualize the Best Server area of the main BTS sector. The method for

Algorithm III.4 Algorithm DrawMap – function $CreateMap$

```
function CreateMap(array points)
    for i = 0 to map.MarkersCount by 1 do
        max_RxLev ← null
        final_color ← null
        marker ← map.Markers[i]
        for j = 0 to marker.PointsCount by 1 do
            point ← marker.Points[j]
            if point.RxLev < max_RxLev then
                max_RxLev = point.RxLev
                final_color = point.Color
            end if
        end for
        map.Markers[i].Color ← final_color;
    end for
end function
```

visualization of Best Servers is described using a following algorithm. In the first step is divided a map raster into an equal parts and set up a grid (see Algorithm III.1).

IV. Conclusion

This article describes how was developed a new algorithm for visualization the Best Server areas. This algorithm improves the basic positioning methods used in mobile networks to determine the location of mobile station. The final result of proposed method is that we are able to provide more accurate location information and visualize it into the street map. Reference database of network parameters for proposed algorithm is made from measurements in GSM network. The result of Best Server visualization in LoMo Map Viewer is a real Best Server area that improves the basic operator's visualization as shown on Fig. 1. In the further research, we plan to focus on larger areas with lower density of BTS. The aim of this effort is to improve an accuracy of the introduced method by using the parameter TA. In addition, we want to check the influence of number of samples and density of monitored routes on accuracy of Best Server visualization.

V. Acknowledgment

This paper was supported by a grant IT Cluster 2009-2012 No. 5.1 SPK01/029 which is funded by the European Fund for Regional Development and Ministry of Industry and Trade.

References

[1] Syed A. Ahson, Mohammad Ilyas, "Location-Based Services Handbook: Applications, Technologies, and Security," CRC Press, 2010, pp. 158-169.
[2] Nel Samama, "Global Positioning: Technologies and Performance (Wiley Survival Guides in Engineering and Science)," Wiley-Interscience, 2008, pp.255-281.
[3] Raymond Steele, Chin-Chun Lee, Peter Gould "GSM, cdmaOne and 3G Systems," 1.edition, Wiley, 2001, pp. 311-357.
[4] Raymond Steele, Lajos L. Hanzo, "Mobile Radio Communications," 2.edition, Wiley-IEEE Press, 1999, pp. 684-701.
[5] Asha Mehrotra, "GSM System Engineering," Artech House, 1997, pp. 97-121.
[6] Krzysztof Wesolowski, "Mobile Communication Systems," West Sussex, England: John Wiley & Sons, 2002, pp.157-165.
[7] Timo Halonen, Javier Romero, Juan Melero, "GSM, GPRS and EDGE Performance Evolution Towards 3G/UMTS," 2.edition, West Sussex, England: John Wiley & Sons, Ltd, 2003, pp. 11, 14, 49.
[8] Jrg Eberspächer, Hans-jörg Vögel, Christian Bettstetter, Christian Hartmann, "GSM-Architecture, Protocols and Service," 3.edition, West Sussex, England: John Wiley & Sons, Ltd, 2002, pp. 58, 108, 139, 212, 234, 264, 266.
[9] Gunnar Heine, "GSM Networks: Protocols, Terminology and Implementation," Artech House, 1998, pp. 11-19.
[10] Siegmund M. Redl, "An Introduction to GSM," Artech House Publishers, 1995, pp. 218-225.
[11] Michel Mouly, Marie-Bernadette Pautet, "The GSM System for Mobile Communications," Telecom Publishing, 1992, pp. 515-525.
[12] ETSI.: GSM 04.08 version 7.21.0 Release 98. ETSI - France, 2003-12, pp. 29-38.
[13] ETSI.: GSM 45.008 version 9.2.0 Release 9. ETSI - France, 2010-04, pp. 101-107.
[14] ETSI.: GSM 08.08 version 8.9.0 Release 99. ETSI - France, 2001-04, pp. 13-19.
[15] ETSI.: GSM 04.31 version 8.18.0 Release 99. ETSI - France, 2007-06, pp. 11-16.
[16] E. Trevisani, A. Vitaletti, "Cell-ID location technique, limits and benefits: an experimental study," Mobile Computing Systems and Applications WMCSA, 2004, pp. 51 -60 [Sixth IEEE Workshop on Mobile Computing Systems and Applications, p. 51, 2004].
[17] Cheung, K.W. , So, H.C. , Ma, W.-K. , Chan, Y.T. , "Received signal strength based mobile positioning via constrained weighted least squares," Acoustics, Speech, and Signal Processing, 2003, pp. 137 - 140. [IEEE International Conference ICASSP '03, p. 51, 2003].

Implementation of OFDM demodulator using software environment System Generator for DSP

Miroslav Bures, Marek Dvorsky
Department of Telecommunications
VSB – Technical University of Ostrava
17. listopadu 15, 708 33 Ostrava, Czech Republic
miroslav.bures.st1@vsb.cz, marek.dvorsky.@vsb.cz

***Abstract*—This paper deals with software implementation of OFDM demodulator using a software tool *System Generator for DSP* from Xilinx. "System Generator for DSP" is in conjunction on environment MATLAB-Simulink capable of simulating the proposed hardware structures that is synthesized and implemented by the programmable elements in *Field-programmable Gate Arrays* (FPGAs).**

I. Introduction

Nowadays, new digital technologies bringing a significant increase in the quality of transmission information systems. This is particularly the truth in connection with Digital Video (DVB) and Audio (DAB, DRM) Broadcasting systems. All these systems use a digital modulation technique called Orthogonal Frequency Division Multiplexing (OFDM).

System OFDM uses a principle of frequency divided transmission channel. This division is made by hundreds of subcarrier frequencies, which are further modulated by a multilevel modulation such as Quadrature Phase Shift Keying (QPSK) or M-Quadrature Amplitude Modulation (M-QAM). Thus the modulated signal is highly resistant to Inter-Symbol Interference (ISI) and Inter-Carrier Interferences (ICI), which are often caused by multi-path spread signal and the Doppler Effect.

On the receiving side, it is necessary to ensure a coherent demodulation. The main parameter, on which depends an overall bit error rate, is the precision of synchronization.

II. Digital Radio DRM

Digital Radio Mondiale (DRM) is a universal, worldwide open standard of digital radio, which was designed for low (LF), medium (MF) and High (HF) frequency bands. Currently available specification of DRM+ is designed to work at Very High Frequency (VHF) band.

A. DRM specification

1) Source coding: For a source coding of input audio streams are in DRM used three basic codecs. Differences between codecs are mainly the quality of sound and bit rate. DRM uses a combination of Advanced Audio Coding (AAC), Code Excited Linear Prediction (CELP) and Harmonic Vector Excitation Coding (HVXC). Codec AAC provides the highest audio quality compared to CELP and HVXC, which reach a lower quality and lower bit rate. AAC codec is used for standard audio coding. CELP and HVXC is used especially for encoding spoken word. To increase the efficiency may be used Spectral Band Replication (SBR).

2) Transmission modes: A distorted receiving signal is caused in the radio channel by disturbing noise. The digital system DRM is equipped with mechanisms that ensure errorless signal decoding in order to overcome this signal distortion. Choosing an optimal combination of modulation type and coding rate is the goal how ho to ensure a signal reception in the maximal possible quality. These combinations are called Transmission Modes (Modes of robustness).

DRM system distinguishes four basic transmission modes. These schemes can have a large transfer rate and little robustness or, conversely, the lowest bit rate, but great robustness. A basic overview of transmission modes with the specific features of the OFDM signal is defined in Table I.

TABLE I: Parametrs of OFDM signal

Transmission parameters	Transmission modes			
	A	B	C	D
$\mathrm{T_u}$ [ms]	21,00	21,33	14,66	9,33
$\mathrm{T_g}$ [ms]	2,66	5,33	5,33	7,33
$\mathrm{T_g/T_u}$ [-]	1/9	1/4	4/11	11/14
$\mathrm{T_s = T_u + T_g}$ [ms]	26,66	26,66	20,00	16,66
Δf [Hz]	$41^{2/3}$	$46^{7/8}$	$68^{2/11}$	$107^{1/7}$

Parameter T_s specifies the duration of the OFDM symbol, T_g is duration of a Guard Interval (GI) and T_u is the duration of useful (data) part of the OFDM symbol, i.e. without GI. Δf is the distance between two sub-carriers waves and is given by following equation:

$$\Delta f = \frac{1}{T_u} \qquad [Hz] \tag{1}$$

where T_u is the duration of the useful OFDM symbol part.

A broadcast signal is organized into Super Frames blocks. One Super Frame consists of a three normal frames. Each frame contains of N_s OFDM symbols with symbol duration T_f. OFDM symbol contains data or synchronization information for the demodulation and decoding. The frame thus consists of the data, pilot and control signals.

For the purpose of OFDM demodulation are substantial pilot signals that carry information about the frequency, time and frames synchronization. The receiver based on the knowledge

of pilot signals is capable of carrying out equalization transmission channel and coherent demodulation. Pilot signals are evenly distributed across time and frequency domains.

III. Theoretical design OFDM demodulator in the DRM system

Internal structure of DRM receiver can be divided into several basic logical blocks, which include OFDM demodulator. The other units are the radio-frequency circuits, frame decoder, audio and data flow decoders. A simple block diagram can be seen in Figure 1.

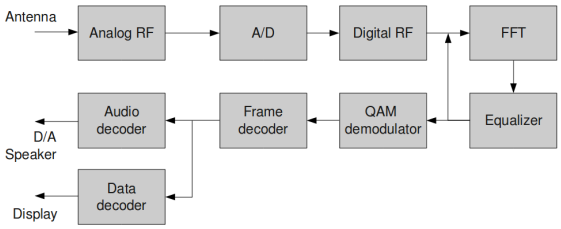


Fig. 1: Block circuit of DRM receiver

For the subsequent implementation was chosen a block of Fast Fourier Transformation (FFT), and 16-QAM demodulator. Furthermore, the functionality of these blocks has been verified by simulations and then their implementation into FPGA.

IV. Implementation of selected blocks in the environment of the System Generator for DSP

A. Block of Fast Fourier Transformation

In an environment of System Generator for DSP is available the block Fast Fourier Transform v7.1. This block implements the Cooley-Tule algorithm for calculating the Discreet FT (DFT). This operation is defined as:

$$X_k = \sum_{n=0}^{N-1} x_n e^{-\frac{2\pi i}{N} kn} \quad k = 0, \ldots, N-1 \qquad (2)$$

Block FFT v7.1 has three main inputs. Inputs *xn_re* and *xn_im* are complex components of the input vector and input *start* determines the beginning of the FFT frames. The outputs *xk_re* and *xk_im* are then the calculated values of the FFT. Output *dv* is a validity indicator of the output data and output *done*, then of H level inform about completion of the FFT calculation of the current frame.

Among basic settings block belongs, in particular, the FFT window size in the range 8 - 65 536. Another setting is the choice of implementation algorithm, i.e. whether block will be implemented using pipelined algorithm or Radix-4/Radix-2. For more information on this setting can be found in [4]. Timing diagram of the simulation is on Figure 3, where can be seen the input and output signals.

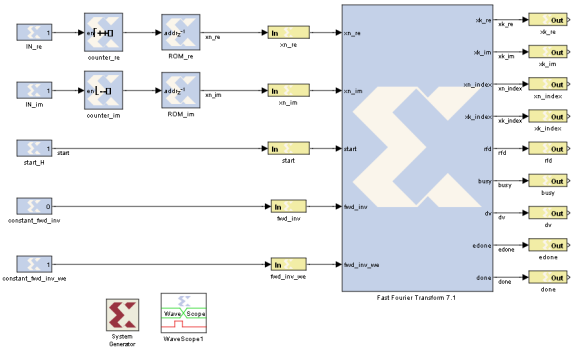


Fig. 2: Simulation diagram of FFT blok

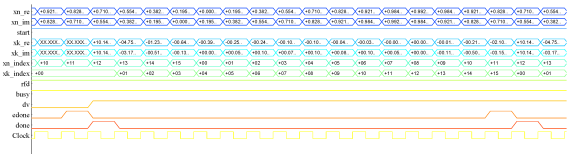


Fig. 3: Time diagram simulation of FFT block

B. QAM demodulator

Quadrature Amplitude Modulation is a multi-level modulation when the modulation signal affects the amplitude and phase of a carrier wave. This provides high spectral efficiency.

The modulation states of M-QAM modulation are composed of multiple n-bits groups which describe modulation symbols, where $M = 2^n$. Placement all the symbols then form the constellation diagram, see Figure 4 These symbols are then transmitted in the form of signal elements which have a duration that equal to a symbol's period T_s.

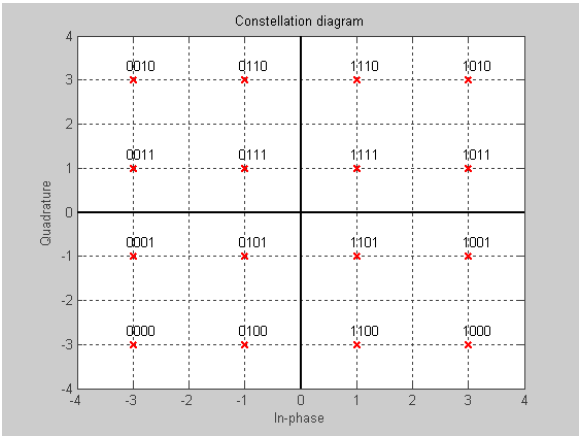


Fig. 4: Constellation diagram for 16-QAM

M-QAM modulation is usually carried out by quadrature modulators also known as vector's (IQ) modulator. A function of a quadrature modulator is based on the fact that any signal of constant ω_c angular frequency and arbitrary time variable phase $\Phi_c(t)$ and any time variable amplitude $U_c(t)$ expressed in usual form:

$$u(t) = U_c(t) \sin\left[\omega_c t + \Phi_c(t)\right] \qquad (3)$$

The same signal is also possible to compose of two components with the same frequency and with a constant mutual phase shift of 90degree (know as Quadrature Component) and the amplitude I (t), Q (t), i.e.

$$u(t) = I(t)\sin(\omega_c t) + Q(t)\cos(\omega_c t) \quad (4)$$

M-QAM signals can be demodulated using quadrature demodulator, which function is essentially inverse function to the modulator [5].

TABLE II: Partition of QAM symbols on di-bits

I,Q component	$a_0 a_1$
3	10
1	11
-1	01
-3	00

The diagram in Figure 5 implements a 16-QAM demodulator shown in Table II. Bit a_0 is generated based on the marks received sample. Multiplexer makes an absolute value of the sample, which ensures that further processing is in positive values.

Other blocks of proposed structure realize the decision the value of bit a_1. Subsequently di-bit I and Q components composed into a serial data output in meaning of a demodulated data.

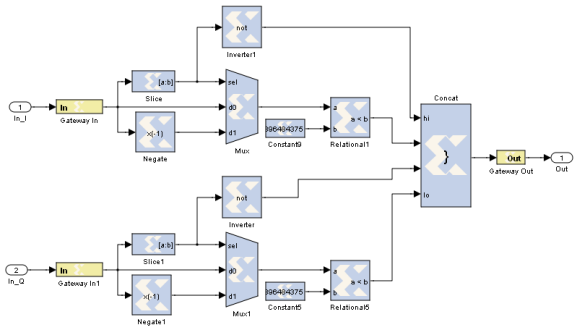


Fig. 5: Block diagram of 16-QAM demodulator

Figure 6 shows the block of simulation involvement. Randomly generated input vector is mapped by QAM modulator and then transmitted over the AWGN channel with the Signal to Noise Ratio (SNR) 10 dB. Demodulated signal is then compared with the broadcast signal and the calculated value of BER.

V. Implementation of selected blocks

After simulation and functionality verification was generated Verilong Hardware Description Language (VHDL) code for both units and then synthesized and implemented into Xilinx ISE development environment. For implementation was chosen chip from product line Spartan3E (xc3s250e in SMD package TQFP144 and speed grade -5).

After synthesis and subsequent mapping of the hardware structure (this is known as the Place&Route) found the overall

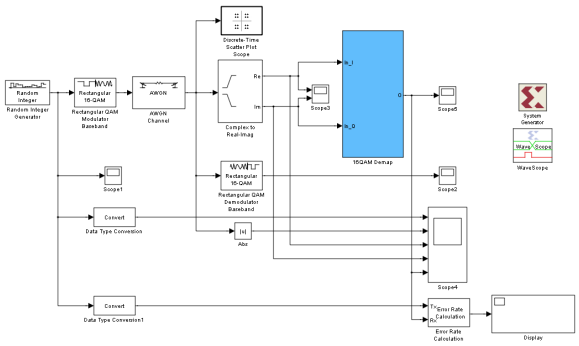


Fig. 6: Simulation diagram of 16-QAM demodulator

TABLE III: Utilization circuit with block FFT

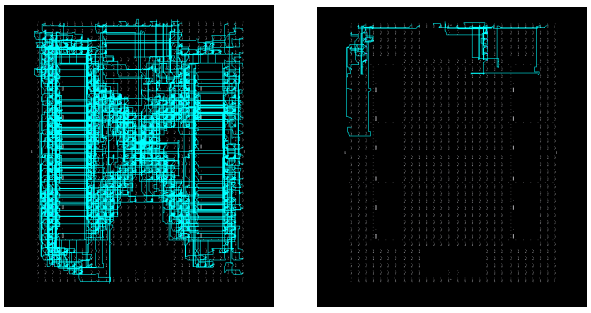
Structure	Used	Total	Percent
External IOBs	59	108	54 %
4 input LUTs	829	4.896	16 %
Slices	730	2.448	29 %
SLICEMs	167	1.224	13 %
BUFGMUXs	1	24	4 %
MULT18X18SIOs	3	12	25 %
RAMB16s	1	12	8 %

TABLE IV: Utilization circuit with block QAM demodulator

Structure	Used	Total	Percent
External IOBs	36	108	33 %
4 input LUTs	38	4.896	1 %
Slices	20	2.448	1 %
SLICEMs	0	1.224	0 %
BUFGMUXs	0	24	0 %
MULT18X18SIOs	0	12	0 %
RAMB16s	0	12	0 %

use of the circuit. These parameters are shown in Tables III and IV

Figure 7 shows the core of the circuit with the placement and interconnection of the internal structures.



(a) FFT (b) 16-QAM

Fig. 7: Layout of blocks on a chip

VI. Conclusion

The aim of this article is to determine the suitability of software "System Generator for DSP design" and implementation of OFDM demodulator blocks. System Design using MATLAB-Simulink and then generating the VHDL code is very efficient. The advantage is especially high efficiency of the proposal in terms of total time spent implementing and debugging the resulting aplication. Another significant advantage is verification of the circuit design by simulation using standard components. Using this kind of implementation brings advanced development of hardware receivers of digital radio into common practice. The new development will provide a massive series production and the price reduction, which is currently now very high.

The practical implementation verified results of the theoretical simulation. Followed measurements have confirmed the findings of simulations and theoretical assumptions. Detailed test results and comparing the simulated 64-QAM demodulator will be republished. In the further work, we are going to focus on implementation of other parts of the OFDM demodulator such as block of synchronization and equalization. It will be also followed by testing of the OFDM demodulator.

References

[1] ETSI ES 201 980 V2.1.1, *Digital Radio Mondiale (DRM); System Specification*, France: European Telecommunications Standards Institute, 2004.

[2] CHEN, Chen, et al., *Synchronization Acquisition Methods for DRM Systems*, In Vehicular Technology Conference, 2006, Montreal: IEEE Xplore, 2006, ISBN 1-4244-0062-7.

[3] KELLER, Thomas, et al., *Orthogonal Frequency Division Multiplex Synchronization Techniques for Frequency-Selective Fading Channels*, In IEEE Journal on Selected Areas in Communications, IEEE Xplore, 2001.

[4] Xilinx, *System Generator for DSP: Reference guide (v10.1)*, www.xilinx.com [online], 2008.

[5] ŽALUD, Václav, *Moderní radioelektronika*, 1st ed. Praha: BEN - technická literatura, 2000, ISBN 80-86056-47-3.

Measurement and sensors, Biomedical Engineering, Multimedia

chairman:

Alparslan ILGIN and Marek NERUDA

Fast DCT Video Compositing for Multi-Point Video Conferencing

Hakkı Alparslan Ilgın and Hakkı Gökhan İlk
Dept. of Electronics Engineering
Ankara University
Ankara, Turkey
ilgin@eng.ankara.edu.tr,
ilk@eng.ankara.edu.tr

Miroslav Voznak
Dept. of Telecommunications
Faculty of El. Eng. and Computer Science
VSB – Tech. Univ. of Ostrava
Ostrava, Czech Republic
miroslav.voznak@vsb.cz

Luis F. Chaparro
University of Pittsburgh
Pittsburgh,
PA 15261, USA
chaparro@ee.pitt.edu

Abstract—**A Discrete Cosine Transform (DCT)-based video compositing for multi-point video conferencing with a fast DCT method for integer and rational image resizing is presented. Regular quantization scheme used in common video coding standards is replaced by a significance map coding in order to improve the quality of the composited videos. In a typical video compositingsystem, video sequences coming from different points are composited into one video streamand sent to receivers using a single channel. This compositing typically requires decoding,decimation and re-encoding.An efficient technique is to perform the whole process in the DCT domain resulting in speed improvement by simply avoiding comprehensive operations such as inverse and forward DCT. To comply with different standards, compositing several video sources inmulti-point video conferencing requires rationaldecimation as well as integer image resizing whichwe address in this paper. Our procedure is illustrated and compared with other methods by means of simulations withdifferent video sequences.**

Keywords-Multi-point video conferencing; video compositing; image resizing; DCT decimation; DCT interpolation; significance map coding; zerotree coding.

I. INTRODUCTION

For multi-point video conferencing [6], compositing of different video sequences into one video sequenceat a node point consists of decimation of incoming videos, combining of the decimated videos and encoding of the combined video. Inverse and forward Discrete Cosine Transform (DCT) and recalculation of motion estimation and compensation are the most computationally expensive operations of the whole process. Therefore a very efficient approach is to perform the compositing in the compressed (DCT) domain rather than in the spatial domain. Video compositing methods using DCT domain decimation and transcoders are more efficient than methods using spatial domain decimation and hybrid codecs [7, 8]. In this paper, we consider DCT compositing approach, develop a general scheme for fast image resizing, previously considered in [11] and [13], and use a significance map coding, which became popular with embedded zerotree wavelet encoding [1] and further improved in [4] called SPIHT, to encode the DCT coefficients obtained from the composited video frames.

For image resizing by a factor of N, we first transform an NxN array of 8x8 DCT blocks (commonly used in video compression standards such as H.263xx, H.264 and MPEGs) into one DCT matrix of size 8Nx8N. Transformed matrix is then masked to obtain the low-frequency DCT coefficients which represent the decimated video frame. Transformation of an array of DCT blocks into one DCT block was introduced in [14]. When looking closely at the transformation matrices very sparse matrices can be found, allowing very fast transformation to be defined. Similar characteristics were first discussed in [13] for decimation by 2. In this paper we show the general case. Proposed algorithms allow fast decimation for integer or rational factors. When the decimation factor is rational, an additional transformation for the masked array is needed. The efficiency of the algorithm is increased by considering only the low-frequency DCT coefficients of the blocks to be decimated. Typically, quantized DCT blocks display zero or very small coefficients corresponding to higher-frequencies.

To expedite the compositing, we use motion vector information from the input streams to estimate -instead of to compute- the motion vectors of the composited video. Likewise, by means of significance map encoder we adapt the quantizer to improve the image quality for a desired bit rate, thus attaining bit rate control.

The rest of the paper is organized as follows. In section 2, we briefly review the DCT compositing system. In section 3 we introduce a fast way to decimate and composite video frames in the DCT domain. In section 4, significance map coding of DCT coefficients and its advantage of using it in the compositing structure is discussed. Simulation results and conclusions are presented in the last two sections.

II. DCT COMPOSITING

In DCT-based video compositing for multi-point video conferencing [6-8], video streams from different sources are decoded by DCT domain transcoders, and then decimated and composited into one video stream as indicated in Fig. 1. Transcoders convert incoming video streams into consecutive DCT images, thus reducing the processing time since there is no inverse DCT operation in the DCT transcoder -the opposite

of a conventional hybrid decoder, which requires inverse DCT to reconstruct the video frames in the spatial domain.

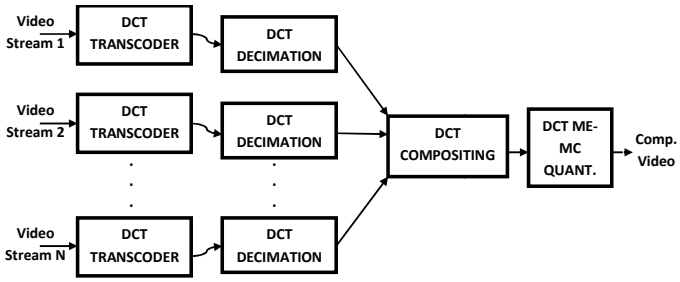


Figure 1. Transcoding and compositing in the DCT domain

The necessity of keeping all operations in the DCT domain demands computing new best matched or optimal DCT blocks for motion compensation. Since the frame being encoded is divided into 8x8 DCT blocks, the optimal block for motion compensation may not coincide with previously computed DCT blocks. Therefore it needs to be obtained using neighboring DCT blocks. Three overlapping cases for an optimal block can be considered: (1) overlapping with one block, (2) overlapping with two vertical or horizontal blocks, and (3) overlapping with four blocks as shown in Fig 2. One way to obtain an optimal DCT block, assuming it does not coincide with an existing block, is to find the inverse DCT of the overlapping DCT blocks with the optimal one and then compute forward DCT of the optimal block from the spatial domain pixels. However, this processing does not remain fully in the compressed domain and increases the number of operations [7].

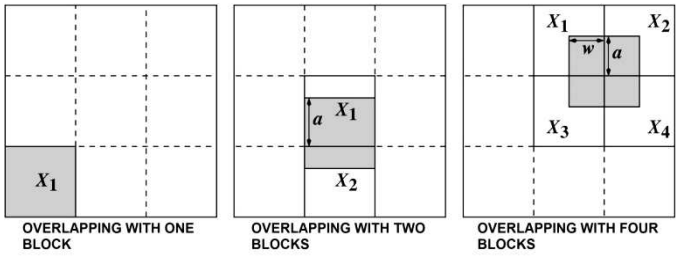


Figure 2. Overlapping cases for an optimal block

To decrease the computational complexity and thus to remain in the DCT domain, DCT windowing and shifting matrices are used to obtain optimal DCT blocks in [7]. Using these matrices improves the speed of the DCT motion compensation [7, 9]. However sparser DCT windowing and shifting matrices as suggested by [10, 11] requires less operations. Additionally the more the first case is met the faster the transcoding is implemented. In this case, motion vectors are either zero or integer multipliers of a block size, meaning that no computation is required to obtain optimal block. As explained later, by decreasing the decimation/interpolation complexity using the proposed algorithms we achieve more speed improvement for overall operations in the DCT compositing system. Some examples illustrating decoding time improvements by regular [7] and fast [11] DCT transcoders are shown in Table I. We use fast transcoders [11] to obtain reconstructed DCT frames. Decimation is done with the proposed computationally efficient algorithms obtaining decimated DCT frames. These frames are then put together to construct composited DCT frames. The last block in Fig. 1 represents the encoding process of the composited video. Here the absence of the forward DCT operation, a must in hybrid-domain compositing systems, is yet another improvement of the DCT approach [6]. We estimate -rather than compute - the motion vectors associated with the blocks of the composited video using motion vectors of the input videos resulting in lower complexity motion estimation [12]. The quantization scheme consists in significance map coding the DCT coefficients, which is explained in Section 4.

TABLE I. AVERAGE SPEED IMPROVEMENTS BY DCT TRANSCODER AND FAST DCT TRANSCODER OVER HYBRID DECODER

Video Sequence	**Average Speedup (%)**	
	DCT Transcoder	***Fast DCT Transcoder***
Claire	22.8	47.6
Salesman	33.3	43.3
Hall	33.7	43.1

III. DCT BLOCK TRANSFORMATION

Many of the video compositing standards require decimation by an integer factor, but some also require decimation by a rational factor to get the mixed views as shown in Fig. 3. To keep the processing in the compressed domain, it is thus necessary to have efficient decimation and interpolation methods. The first issue in DCT decimation is the transformation of an array of DCT blocks into one block. Suppose we have an array of *N*x*N* DCT blocks, where the integer $N \geq 2$ is the decimation factor, each of size 8x8, and we wish to transform it into one DCT block of dimension 8*N*x8*N*. This would correspond to finding the inverse DCT of each of the 8x8 blocks, arranging them into one array of dimension 8*N*x8*N*, and finally finding its 8*N* two-dimensional DCT. A more efficient approach was proposed in [14], which as we show can be attained by means of an orthonormal matrix transformation. The problem is to find an orthonormal transformation matrix T^{8N}, i.e., $(T^{8N})^t T^{8N} = I^{8N}$, where I^{8N} is 8*N*x8*N* identity matrix and *t* is the matrix transport operators, and such that

$$X^{8N} = T^{8N} \begin{bmatrix} X_{11}^{8} & \cdots & X_{1N}^{8} \\ \vdots & \ddots & \vdots \\ X_{N1}^{8} & \cdots & X_{NN}^{8} \end{bmatrix} (T^{8N})^t \qquad (1)$$

where $\{X_{ij}^{8}\}_{i,j=1,\ldots,N}$ and X^{8N} are 8x8 and 8*N*x8*N* two-dimensional DCT blocksrespectively. By definition, if the *M*x*M* DCT operation matrix is

$$S^M = \{s(k,n)\}_{k,n=0}^{M-1} = \{0.5c(k)\cos(\tfrac{(2n+1)k\pi}{2M})\} \qquad (2)$$

the two-dimensional DCT of the spatial domain matrix $x = \{x(n,m)\}_{n,m=0}^{M-1}$ is given by

$$X = \{X(k,l)\}_{k,l=0}^{M-1} = S^M x (S^M)^t \qquad (3)$$

Figure 3. Composited frames; five with decimation factor N=3, and one with decimation factor N=2/3

and the inverse DCT is given as

$$x = (S^M)^t X S^M. \tag{4}$$

The transformation can be easily obtained by finding the DCT of the identity matrix I^{8N}(each $X_{ii}^8 = I^8$and $X_{ij}^8 = 0, i \neq j$)using the forward DCT definition in (3) and orthonormality property of S^M so that

$$T^{8N} = S^{8N}\text{diag}[(S^8)^t]. \tag{5}$$

To obtain the forward block transformation, let the 8Nx8N matrices S^{8N} and T^{8N} be expressed as follows

$$S^{8N} = [S_1^{8N} S_2^{8N} \cdots S_N^{8N}]$$

$$T^{8N} = [T_1^{8N} T_2^{8N} \cdots T_N^{8N}], \tag{6}$$

where the sub-blocks

$$\{T_i^{8N} = S_i^{8N}(S^8)^t\} \tag{7}$$

are of dimension8Nx8. Thus the representation of X^{8N} in terms of the X_{ij}^8,or the forward block transformation, is

$$X^{8N} = \sum_{i=1}^N \sum_{j=1}^N T_i^{8N} X_{ij}^8 (T_j^{8N})^t. \tag{8}$$

Therefore 8Nx8$N$$X^{8N}$ DCT block is obtained directly from an NxN array of 8x8 DCT blocks.The decimation can now be obtained by masking the resulting block to obtain the more significant DCT coefficients.The inverse transformation of each ofthe 8x8 DCT blocks is given by

$$X_{ij}^8 = (T_i^{8N})^t X^{8N} T_j^{8N} \tag{9}$$

using the orthonormality of the DCT transformation. This will allow us to perform interpolation.

A. Fast DCT Decimation

To simplify the notation consider decimation by 2 (N=2). The equations for other cases for N>2 can be easily extended.Applying the direct transformation, we obtain a 16x16 DCT block from four 8x8 given DCT blocks.Masking the X^{16} matrix to extract the low-frequency coefficients we obtain the decimated DCT array as

$$X_d = [I^8\ 0]X^{16}[I^8\ 0]^t$$

$$= \sum_{i=1}^2 \sum_{j=1}^2 F_i^8 X_{ij}^8 (F_j^8)^t \tag{10}$$

where $\{F_i^8 = [I^8\ 0]T_i^{16}\}_{i=1,2}$ are of size 8x8.The above equation can be efficiently implemented using the sum anddifference representations of$\{T_i^{16}\}$;

$$D_1^{16} = 0.5(T_1^{16} + T_2^{16})$$

$$D_2^{16} = 0.5(T_1^{16} - T_2^{16}) \tag{11}$$

where D_1^{16} and D_2^{16} are 16x8 transformation matrices. Replacing the $\{T_i^{16}\}$ in terms of $\{D_i^{16}\}$ matrices in thedirect transformation we have

$$X^{16} = [Y + Z](D_1^{16})^t + [Y - Z](D_2^{16})^t$$

$$Y = D_1^{16}(X_{11}^8 + X_{21}^8) + D_2^{16}(X_{11}^8 - X_{21}^8)$$

$$Z = D_1^{16}(X_{12}^8 + X_{22}^8) + D_2^{16}(X_{12}^8 - X_{22}^8) \tag{12}$$

Transformation matrices $\{D_i^{16}\}$ are sparser than $\{T_i^{16}\}$,so the forward block transformationin(12) is more efficient than in (8).Therefore decimated DCT block $X_d$is efficiently obtained as

$$X_d = [Y_d + Z_d](A_1^8)^t + [Y_d - Z_d](A_2^8)^t$$

$$Y_d = A_1^8(X_{11}^8 + X_{21}^8) + A_2^8(X_{11}^8 - X_{21}^8)$$

$$Z_d = A_1^8(X_{12}^8 + X_{22}^8) + A_2^8(X_{12}^8 - X_{22}^8) \tag{13}$$

where the $\{A_i^8 = [I^8\ 0]D_i^{16}\}_{i=1,2}$ matrices of size 8x8are sparser than the$\{F_i^8\}$ matrices defined in (10), and make the decimation faster.

A faster algorithm can be obtained by observing that high frequency coefficients in a DCT block aretypically zero or very small, so that when set to zero inverse DCT is not very different from the original spatial domain block. Considerthat the DCT blocks to be decimated have qxq ($1 \leq q \leq 8$) low-frequency nonzero components and we set the rest to zero, i.e.,

$$X_{ij}^8 = \begin{bmatrix} X_{ij}^q & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} I^q \\ 0 \end{bmatrix} X_{ij}^q [I^q\ 0] \tag{14}$$

Replacing these blocks in (10) gives

$$X_d^q = \sum_{i=1}^{2}\sum_{j=1}^{2} G_i^8 X_{ij}^q (G_j^8)^t \quad (15)$$

where X_d^q is 8x8 decimated DCT block and $\{G_i^8 = F_i^8[I^q\ 0]^t\}_{i=1,2}$, $1 \le q \le 8$. $\{G_i^8\}$ matrices of size 8x$q$clearly have fewer entries than$\{F_i^8\}$.However we have the decimation in a faster way if we use the decimation in (13) as

$$X_d^q = \left[Y_d^q + Z_d^q\right](B_1^q)^t + \left[Y_d^q - Z_d^q\right](B_2^q)^t$$

$$Y_d^q = B_1^q\left(X_{11}^q + X_{21}^q\right) + B_2^q\left(X_{11}^q - X_{21}^q\right)$$

$$Z_d^q = B_1^q\left(X_{12}^q + X_{22}^q\right) + B_2^q\left(X_{12}^q - X_{22}^q\right) \quad (16)$$

where $\{B_i^q = A_i^8[I^q\ 0]^t\}_{i=1,2}$ .Obviously, $\{B_i^q\}$ matrices have fewer entries than $\{F_i^8\}$ and also $\{A_i^8\}$. The proposed DCT block transformation and decimation for N=2 is shown in Fig. 4.

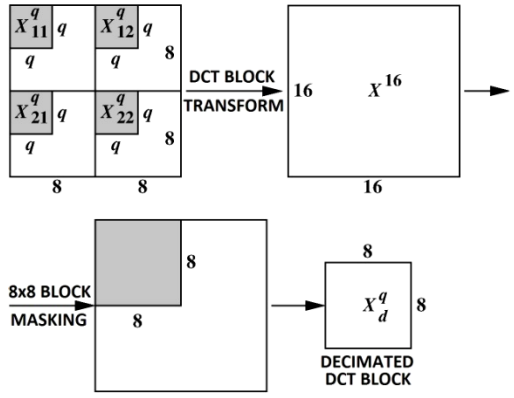


Figure 4. Proposed integer DCT decimation for decimation factor N=2

When compared to other methods, our procedure (q=4) has the lowest complexity(measured in terms of multiplications and additions per pixel) [17] as shown in Table II.

TABLE II. COMPUTATIONAL COMPLEXITY OF DECIMATION METHODS FOR N=2

Method	**Multipl./Pixel**	**Add./Pixel**
Prop. (q=4)	1.25	1.25
[13]	1.25	1.25
[16]	1.00	2.00
[15]	1.89	2.06
[7]	4.00	4.75
Spatial	3.44	9.82

We also showed the effectiveness of the proposed method in using low bit-rate single stream video coding [17].Since the decimation in [13] has the same number of operations we compare it with our method in terms of Peak Signal to Noise Ratio (PSNR). The results are shown inTable III. As expected in our method, the larger the q is (i.e., $4 \le q \le 8$), the better the interpolation (as seenin Table III, but the more complex the implementation. In Table III, our method with first$q$=4 has the same number of operations for both decimation and interpolation as in[13]. As seen PSNR results are very close. Additionally, with a slight increase in the number of operations in the interpolation while keeping thecomplexity of the decimation as before, the PSNR values (in dB) of the interpolated frames are improved (see q=4 (ii) in the table).

TABLE III. PSNR COMPARISONS OF THE PROPOSED METHOD WITH [13] FOR N=2

Video Frame	**[13]**	***q*=4 (i)**	***q*=4 (ii)**	***q*=8**
Miss America	38.97	38.88	39.13	39.43
Foreman	32.74	32.58	32.82	33.19
Hall	28.64	28.48	28.76	29.17

Decimation by an integer, where N>2, proceeds in a very similar way to decimation by N=2. In general, for an integer decimation factor N, an array of NxN DCTs of size 8x8 is transformed into one array of size 8Nx8N, and then it is masked to obtain the decimated 8x8 DCT. For example for N=3, a 24x24 DCT matrix is obtained and just as before each 8x8 block is represented with qxq coefficients (q< 8) and the rest zero for faster implementation. Computational complexity depends on the value of q, and PSNR is better for larger values of q. However decimation factors, such as N=3 or N=4, for even small values of q the quality of the interpolated images does not differ visually or in terms of PSNR as shown in Table IV. For example, for N=4, the PSNR results are the same for q=4, 5, 6 and 8, and are very close to the PSNR for q=2. Thus only the 2x2 low-frequency parts of the 8x8 DCT blocks may be sufficient to decimate DCT blocks. This constitutes a significant computational saving.

TABLE IV. PSNR COMPARISONS OF A CLAIRE FRAME FOR DIFFERENT DECIMATION FACTORS WITH SEVERAL DCT BLOCK SIZES

	Method					
N	q=2	q=3	q=4	q=5	q=6	q=8
2	28.99	30.62	33.93	34.10	34.16	34.16
3	28.90	30.12	30.32	30.33	30.32	30.33
4	28.84	29.25	29.30	29.30	29.30	29.30
2/3	28.66	30.30	34.10	35.04	37.07	37.08

In Fig. 5, PSNR results of decimated/interpolated (N=2, $1 \le q \le 8$) frames are shown. When q=1 only the DC values are used to obtain the decimated frames and it is still possible to see a coarse reconstructed image. For the case of N=2, q=4 is the most appropriate value, since there is no significant PSNR improvement for q>4. This shows the efficiency of the proposed method which uses smaller portions of DCT blocks efficiently. It is also possible to decimate by a rational number, e.g., N=2/3 or N=3/4. In this case the algorithm requires additional computations. When N=2/3, we first need to transform a 3x3 array of 8x8 DCT blocks into one of size 24x24. The masking to get 2/3 gives us a DCT block of dimension 16x16, that needs to be converted into four 8x8

blocks finally. These additional computations increase the complexity of the decimation for rational factors.

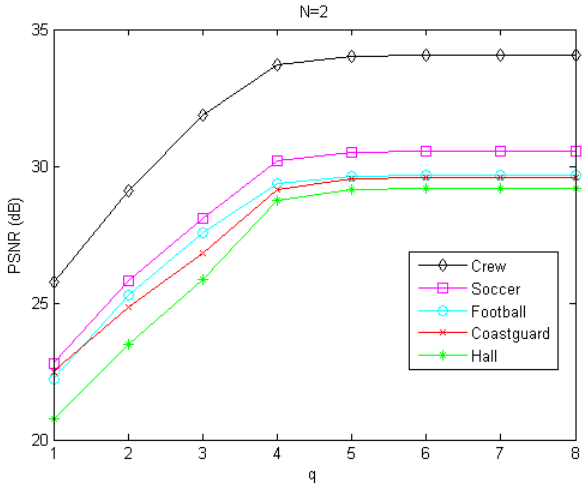


Figure 5. Decimation by 2 using different *q* values

As in the integer case, only *qxq* parts of DCT blocks need to be used to obtain 24x24 DCT block as shown in Fig. 6. This substantially reduces the number of operations. For instance, if all coefficients of the 8x8 DCT blocks are used, which corresponds to the case where *q*=8, the number of multiplications and additions per pixel to obtain the final four 8x8 DCT blocks are 21.58 and 20.81 respectively. If only 6x6 portions of the input DCT blocks are used (*q*=6), operations reduce to 18.38 multiplications and 17.16 additions per pixel. In terms of PSNR, there is almost no difference between *q*=6 and *q*=8 cases as shown in Table IV.

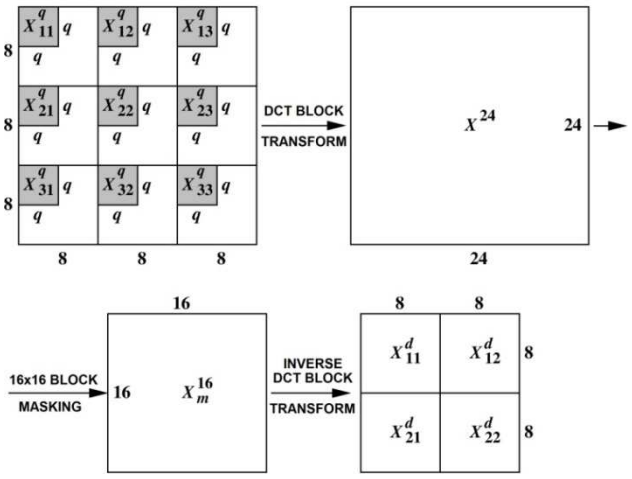


Figure 6. Proposedrational DCT decimation (*N*=2/3)

IV. SIGNIFICANCE MAP CODING

To encode the composited video, we replace regular scalar quantizing with the DCT significance map based zerotree coding (DCT-EZT) which is previously used for wavelet [1] and then for DCT [2, 3]. In [4] compression ratio is improved. It has also efficient differential approach in [5]. To encode the DCT coefficients by a zerotree coder they are rearranged into a hierarchical subband structure exploiting the dependency between them. In this paper we arrange video frames consisting of 8x8 DCT blocks into a 3-scale hierarchical structure resulting in ten subbands as shown in Fig. 7. The highest subband, LL_3, includes all DC coefficients. All other subbands include AC coefficients. Since the structure is the same as the wavelet coefficients HL_i, LH_i and HH_i includes horizontal-, vertical- and diagonal-detail AC-DCT coefficients. DC coefficients contain most of the energy of the frame. Therefore the quality of the decoded image depends in great part on the DC coefficients. Thus DC coefficients are firstly encoded and then the low to high frequency AC coefficients follow.

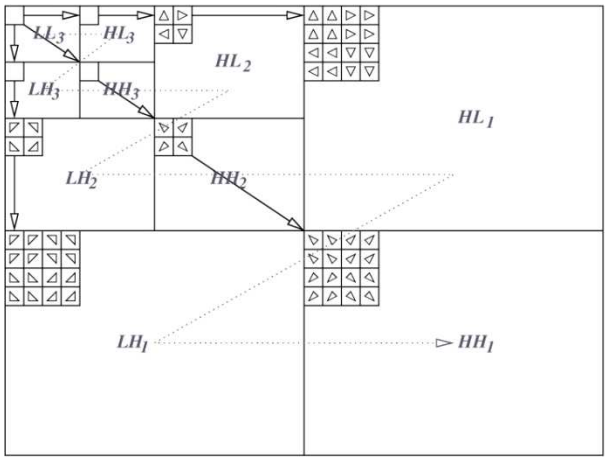


Figure 7. Relationship of arranged DCT coefficients in 3-scale subband structure

Zerotree coding starts with measuring the significance of all coefficients by comparing them with an initial threshold

$$Th_0 = 2^{\lfloor \log_2(\max(|X_{k,m}|)/2 \rfloor} \quad (19)$$

where$\{X_{k,m}\}$are DCT coefficients of a frame of size *K*x*M*. If the magnitude of a coefficient is greater than the threshold, it is considered significant. The advantage of this type of coding is that if a group of coefficients related to each other as tree in differentsubbands is found to be insignificant, only one symbol is sent to decoder stating that the tree has coefficients lower than the threshold called zerotree. This contributes compression ratio. After comparing all coefficients with threshold, a subordinate scan is performed to givemore precision to significant coefficients [1-4]. Threshold is halved aftersubordinate scan finishes. As soon as a symbol is outputted by the zerotree coder, it is encoded by an adaptive arithmetic encoder to obtainembedded bit stream and more compression [1]. A detailed explanation of arithmetic coding can be found in [18]. Process continues until the desired bitrate is reached.

V. SIMULATIONS

We consider three different compositing methods in the simulations: (1) decimation by *N*=2, (2) decimation by *N*=3 and *N*=2/3 and (3) decimation by *N*=3. When *N*=2, four video sequences are composited into one video, while in the mixed-view case six video sequences are composited into one: five videos decimated by 3, and one by 2/3. For *N*=3, composited video frames consist of nine subframes. Each video sequence

has an intra-frame and 49 inter-frames in CIF (Common Intermediate Format). PSNR results are given as the average of the PSNR of 50 frames. In all compositing structures we obtain better results by using DCT-EZT coding than by using regular scalar quantizing with different quantization parameters (*QP*) as suggested in different standards such as H.264 as shown in Table V. Also the reconstructed frames are subjectively better in DCT-EZT case in terms of blocking effect especially at low bit rates. It is further possible to improve the quality of the quantization process by using the approach in [5].

TABLE V. AVERAGE PSNR COMPARISONS FOR COMPOSITED VIDEOS

Structure	**Method**	**PSNR (dB)**		
4-subframe	***Regular***	***QP*=5**	***QP*=10**	***QP*=20**
		36.25	32.48	28.25
	Zerotree	38.20	34.77	31.42
6-subframe	***Regular***	***QP*=5**	***QP*=10**	***QP*=20**
		36.61	33.11	29.07
	Zerotree	39.00	35.50	32.78
9-subfame	***Regular***	***QP*=5**	***QP*=10**	***QP*=20**
		35.30	31.40	26.99
	Zerotree	37.16	33.40	29.61

VI. CONCLUSION

In this paper we show that compositing for multi-point video conferencing can be completely and very efficiently done inthe compressed domain, which is important for real time encoding. The decimation in the DCT domain permits us to increase the quality of theimage by decreasing aliasing, and we show that it can be done with our efficient algorithms using only *q*x*q* portion of DCT blocksfor integer or rational decimation factors. The value of *q* can be chosen to minimize the number of operations in decimationand interpolation for both integer and rational decimation factors without affecting the image quality significantly.The significance map coding allows us to choose the bit rate, and performs better in terms of image quality than theconventional DCT compositing system using regular scalar quantization. It is possible to improve the compression rate by implementing recent improvements on significance map coding. Also bit rate allocation can be done at sub-frame level in order to improve the quality of the interested video sequence or objects. Proposed decimationmethod can be used in conjunction with H.264 SVC to obtain different resolutions at spatial scales of the said video standard.

ACKNOWLEDGMENT

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 218086.

REFERENCES

[1] J. M. Shapiro, ``Embedded Image Coding Using Zerotrees of Wavelet Coefficients," IEEE Trans. Signal Proc., Vol. 41, No. 12, pp. 3445-3462, Dec. 1993.

[2] Z. Xiong, O. G. Guleryuz and M. T. Orchard, ``A DCT-Based Embedded Image Coder," IEEE Signal Processing Letters, Vol. 3, No. 11, pp. 289-290, Nov. 1996.

[3] D. M. Monro and G. J. Dickson, ``Zerotree Coding of DCT coefficients," IEEE Intern. Conf. Image Proc., Vol. 2, pp. 625-628, Oct. 1997.

[4] A. Said and W. A. Pearlman, ``A new, fast, and efficient image codec based on set partitioning in hierarchical trees," IEEE Trans. Circ. and Sys. for Vid. Tech., pp. 243-250, June 1996.

[5] Yang Hu, W. A. Pearlman, "Differential-SPIHT for Image Sequence Coding", Proc. IEEE Intl. Conf. Acoustic, Speech, and Signal Processing, p. 894, Dallas, TX, Apr. 2010.

[6] S.-F. Chang and D. G. Messerschmitt,``A new approach to decoding and compositingmotion-compensated DCT-based images," Proc. IEEE Intl. Conf. Acoustic, Speech, andSignal Processing, Vol. 5, pp. 421-424, Minneapolis, MN, Apr. 1993.

[7] S.-F. Chang and D. G. Messerschmitt,``Manipulation and Compositing of MC-DCT Compressed Video," IEEE Journal on Selected Areas in Commun., Vol. 13, No. 1, pp. 1-11, Jan. 1995.

[8] Y. Noguchi, D. G. Messerschmitt and S.-F. Chang,``MPEG Video Compositing in the Compressed Domain,"Proc. IEEE Intl. Symp. Circuits and Systems, Vol. 2, pp. 596-599, May 1996.

[9] M. Song, A. Cai, J.-a. Sun,``Motion estimation in DCT domain," Proc. IEEE Comm. Technology, ICCT'96, Vol. 2, pp. 670-674, May 1996.

[10] N. Merhav and V. Bhaskaran,``A Fast Algorithm for DCT-domain Inverse Motion Compensation,"Proc. IEEE International Conf. Acoustics, Speech, and Signal Processing,Vol. 4, pp. 2307-2310, May 1996.

[11] N. Merhav and V. Bhaskaran,``Fast Algorithms for DCT-domain Image Downsampling and For Inverse Motion Compensation," IEEE Trans. Circuits and Syst. for Video Technology, Vol. 7, No. 3, pp. 468-476, June 1997.

[12] J. Youn and M.-T. Sun,``A fast motion vector composition method for temporal transcoding,"Proc. IEEE Intl. Symp. Circuits and Systems, Vol. 4, pp. 243-246, June 1999.

[13] R. Dugad and N. Ahuja,``A Fast Scheme for Image Size Change in the Compressed Domain," IEEE Trans. Circuits and Syst. for Video Technology, pp.461-474, Apr. 2001.

[14] J. Jian and G. Feng,``The Spatial Relationship of DCT Coefficients Between a Block and its Sub-blocks," IEEE Trans. Signal Proc., pp. 1160-1169, May 2002.

[15] H. Park, Y. Park and S.-K. Oh,``L/M fold image resizing in block-DCT domain using symmetric convolution," IEEE Trans. Image Process., vol. 12, pp. 1016-1034, Sept. 2003.

[16] J. Mukherjee and S. K. Mitra,``Arbitrary resizing of images in DCT space," IEE Proc.-Vis. Image Signal Process., vol. 152, pp. 155-164, Apr. 2005.

[17] H. A. Ilgin and L. F. Chaparro,``Low-Bit Rate Video Coding Using DCT-based Fast Decimation/Interpolation and Embedded Zerotree Coding,"IEEE Trans. Circuits and Syst. for Video Technology, Vol. 17, No. 7, pp. 833-844, July 2007.

[18] I. H. Witten, R. M. Neal and J. G. Cleary,``Arithmetic Coding for Data Compression," Commun. ACM, Vol. 30, No. 6, pp. 520-540, June 1987.

Modeling of textile DC bulk resistance for antenna design with electrically conductive fibres

Marek Neruda, Lukas Vojtech
Department of Telecommunication Engineering
Czech Technical University in Prague
Prague, Czech Republic
nerudmar@fel.cvut.cz, vojtecl@fel.cvut.cz

Abstract—**Development of high frequency applications influences also textile industry with for example wearable antennas. Modeling of these antennas requires knowledge of material parameters of substrates and functional surfaces. This paper focuses on modeling of textile structure with electrically conductive fibres in order to calculate DC bulk resistance. Measurement method is subsequently proposed for model verification.**

Keywords-bulk resistance; conductive fibres; modeling; textile

I. INTRODUCTION

Modern textile materials are used in many technical applications. One of the perspective branches is so-called Smart textile, which is capable to keep electrical current flow or to operate as electrical sensors. Special group of Smart textile materials is called wearable antenna, which is characterized by electrical conductive textile structure designed for substrate or it is functional material designated for a production of planar antenna structures. These antenna structures are designed as a part of clothing, or they can be applicable directly on the human body [1]. Though the most of synthetic materials in the sheet shape can be used for these applications, textile structures provide many advantages [2].

Modeling of antennas and other high frequency structures requires knowledge of basic material parameters of used substrates and functional surfaces. One of the fundamental parameters is electrical conductivity σ, which is generally derived from electrical resistivity ρ and as a consequence from sheet resistance R_S [3] in the case of metal coating. The most of conductive textile materials are composed from non-conductive material, e.g. based on polyesters, and subsequently the woven fabric is electrochemically coated with the aid of electroplating practice procedures.

If the textile material is electrochemically coated, sheet resistance [Ω/□] is a parameter describing a measure of resistance of constant thin film. It is proportional to $1/(t \cdot \sigma)$. Parameter t is a thickness of metal layer, which is coated on the surface of originally non-conductive textile material. The used metal is characterized by conductivity σ.

As a consequence it is possible to measure the conductivity by DC methods [4]. The results of these measurements are often used also for high frequency applications. It leads to certain inaccuracies, but accuracy of measurement is sufficient for many models [5].

If textile materials are woven from electrical conductive fibres, it is completely different condition. Sheet resistance cannot be considered as a parameter based on upper layer of textile and its thickness. Considering a conductivity of used fibres in whole cross section and capacity, it is important to reason bulk conductivity, not the surface one, during measurement by DC methods. This rule has to be fulfilled also in the case of high frequency applications even though skin effect predominates and the whole cross section is not used.

Structure of textile materials composed form conductive fibres can be seen as grid of resistors. The advantage of this point of view rests in using circuit analysis. Therefore it is possible to model material parameter σ and also measure by DC methods.

The rest of the paper is organized as follows: section 2 describes proposal of textile material model. Section 3 introduces verification of equipotential points presumption by simulation model. Section 4 focuses on measurement procedure and calculation of bulk resistance. Section 5 describes measurement results discussion. Conclusion presents section 6.

II. RESISTANCE OF GRID OF RESISTORS

The textile material can be seen from electrical point of view as an electrical circuit composed from connected resistors and battery. Battery represents electrode poles and resistors correspond to structure of fibres. Fibres form regular shapes in textile material. Considering ideal case the shapes are seen as squares. Then every side of the square represents resistor R′ (marked R1 in figures), Fig. 1. The both poles of battery are connected to shorter sides of textile, i.e. to common nodes of outer resistors of equivalent circuit diagram. The poles of battery were chosen with respect to basic setup for measurement [4]. Calculation of this electrical circuit can be

solved by Kirchhoff's circuit laws or by simplifications of equipotential lines or points. These points are significant, because the points with the same potential can be connected or disconnected as necessary. Then the value of resistor between these equipotential points is equal to zero and it can be eliminated. Following this presumption, the electrical circuit is simplified and resultant resistance is calculated from series-parallel connection of resistors, Fig. 2. It can be calculated as:

$$R = \sum_{1}^{12} \frac{R'}{6} = \frac{12R'}{6} = 2R' \tag{1}$$

If a common formula is derived, resultant resistance is described as:

$$R = \sum_{n=1}^{r} \frac{R'}{s}, n, r, s \in \{ N \} \tag{2}$$

where n, r is number of squares in "horizontal" direction, s represents number of squares in "vertical" direction.

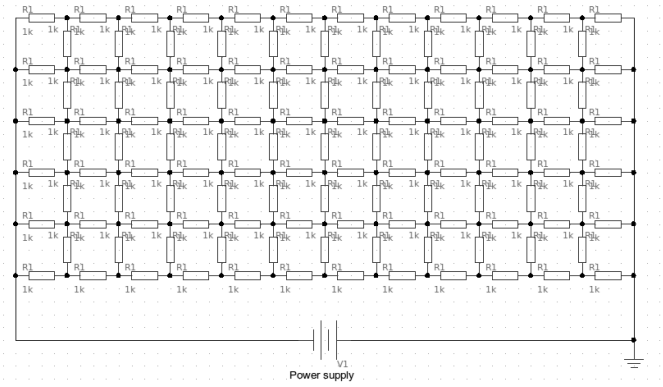


Figure 1. Equivalent circuit diagram of used textile material.

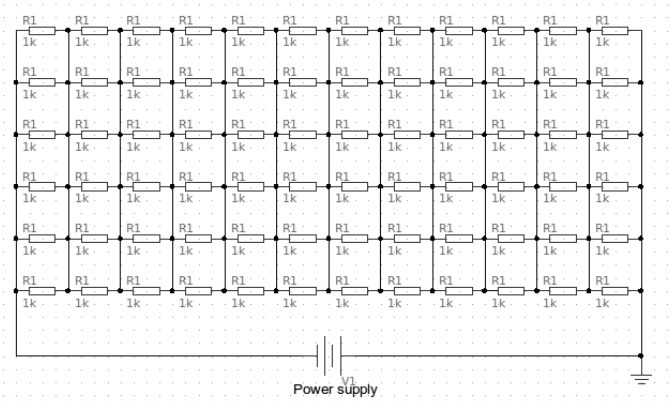


Figure 2. Simplified equivalent electrical circuit with eliminated resistors placed between two points of the same potential.

III. Verification of Equipotential Points Validity

It is possible to use for example software called Oregano, GNOME application, for verifying points with the same potential [7]. Fig. 3 depicts voltage probes, which are placed in five points of different potential. Results are shown in Fig. 4. Different results are naturally obtained by placing voltage probes into the points with the same potential, Fig. 5. Results depicted in Fig. 6 show only one value of voltage for all six measurement points. As a consequence voltage drop between individual resistors is equals to zero, and therefore the resistors can be eliminated. The presumption is thus confirmed and (2) is valid.

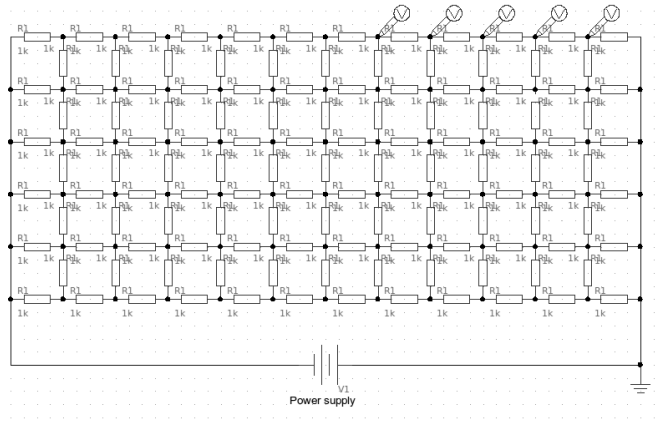


Figure 3. Voltage probes placed in the points with different potential.

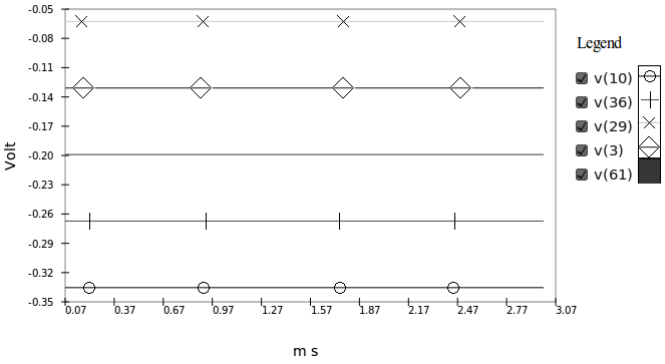


Figure 4. Voltage probes placed in the points with different potential.

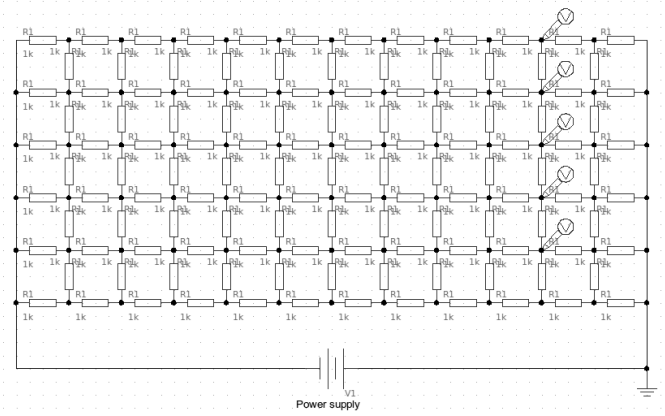


Figure 5. Voltage probes placed in the points with the same potential.

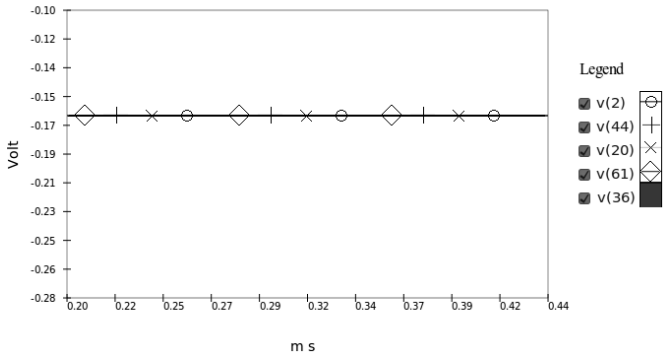


Figure 6. Voltage probes placed in the points with the same potential.

IV. Measurement Procedure of Bulk Resistance

Simplified equivalent electric circuit represents a model of real textile material. Comparison of the model and real characteristics is also possible to perform by measurement method. Resultant resistance of specific textile material requires calculation of *r* and *s* parameters according to (2).

Measurement sample No. 60 is characterized by 25 n/cm warp texture and 20 n/cm weft, where *n* is number of threads. Considering the sample of dimensions 100 x 30 mm, number of threads is calculated as:

$$r' = 10 \cdot 25 = 250 \;\; n \tag{3}$$

$$s' = 3 \cdot 20 = 60 \;\; n \tag{4}$$

As a consequence *r* and *s* parameters are equaled to:

$$r = r' - 1 = 250 - 1 = 249 \;\; n \tag{5}$$

$$s = s' - 1 = 60 - 1 = 59 \;\; n \tag{6}$$

Resultant resistance is then equals to:

$$R = \sum_{n=1}^{r} \frac{R'}{s} = \sum_{n=1}^{249} \frac{R'}{59} = \frac{249R'}{59} = 4{,}22 \cdot R' \tag{7}$$

Resistance R′ of fibre length of one square is obtained from diameter measurement of specific fibre and calculation. It is possible to use basic measurement instruments for approximation of diameter, such as micrometer. Exact value is measured by known methods of textile industry, e.g. microscopic examination, resonance methods or laser diffraction [8]. Resistance value R′ is subsequently calculated as:

$$l_{R'} = \frac{W - s' \cdot d}{s} = \frac{30 - 60 \cdot 0.21}{59} = 0.29 \; mm \tag{8}$$

where $l_{R'}$ is length of *R′*, *s* represents number of resistors in weft, *s′* defines number of threads in weft, *d* is measured diameter and *W* represents width of measured sample.

$$R' = \frac{R_l \cdot l_{R'}}{l_R} = \frac{334.83 - 0.07 \;\; \cdot 0.29}{100} = 0.97\,\Omega \tag{9}$$

where R_l is measured resistance of fiber with length l_R.

$$R = 4.22 \cdot R' = 4.22 \cdot 0.97 = 4.09\,\Omega \tag{10}$$

Measurement diagram is depicted in Fig. 7. The gauging fixture is used for bulk resistance measurement of three textile samples, which are characterized by the same values of *s′* and *s* parameters. Bulk resistance is measured by RLCG bridge ESCORT ELC-3133A with the systematic error s_e=0.07 Ω.

Measurement is not possible to perform by DC power source at high voltage, usually 100 VDC. Resultant resistance is about 1 ohm and it is 100 A at 100 VDC, which can cause overheating and destruction of the sample.

Figure 7. Fibre resistance measurement diagram.

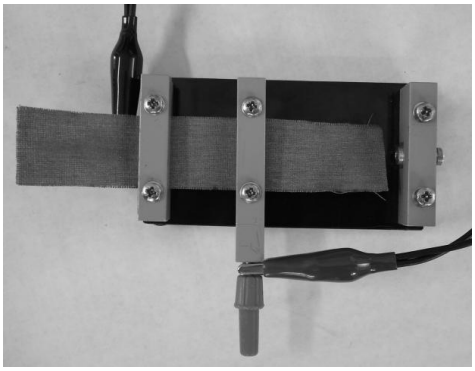


Figure 8. Square setting of electrodes.

Three textile samples are measured in three different positions for two electrode setting, square and rectangular. Square setting measures a sample of dimensions 30x30 mm between electrodes. Rectangular one is 100 mm long, width is also 30 mm. Textile samples are in addition measured with and without insulation material, i.e. paper, which is placed between two through bolts of each electrodes. Parameter R_l is also obtained for different positions of fibre, which is used for all textile samples production. Measurement is performed at 56.8% humidity and 22.9 °C.

TABLE I. MEASUREMENT RESULTS

Measurement setting	**Measurement of Bulk Resistance**		
	No. 60, R [Ω]	***No. 62, R [Ω]***	***No. 64, R [Ω]***
Square, without insulator	1.295 1.473 1.315 avg: 1.361	1.309 1.294 1.235 avg: 1.279	1.269 1.160 1.172 avg: 1.200
Rectangular, without insulator	3.946 3.934 3.977 avg: 3.952	3.834 3.950 3.968 avg: 3.917	3.668 3.671 3.678 avg: 3.672
Square, with insulator	1.358 1.280 1.329 avg: 1.322	1.293 1.249 1.213 avg: 1.252	1.194 1.160 1.197 avg: 1.184
Rectangular, with insulator	3.933 3.984 3.953 avg: 3.957	3.850 3.890 3.977 avg: 3.906	3.644 3.695 3.670 avg: 3.670
One fibre	***R_l [Ω]***		
	331 342 348 395	379 320 277 282	307 358 366 313
	avg: 334.833		

V. RESULTS EVALUATION

Equation (10) gives the bulk resistance result of model R=4.09 Ω for rectangular setting of electrodes. Measurement results for the same setting are almost the same for all measurement, e.g. R=3.952 for No.2 without insulator. It confirms the described model is valid. Inaccuracy is caused by measurement method for diameter determination. Improvement of this method was described.

Influence of measurement error caused by fixation of textile sample is insignificant because bulk resistance takes the same values for measurement with and without insulator. Second reason is results for rectangular setting are about 3.3 multiple of square setting, which is equaled to ratio of their length. For example: No. 64, without insulator, R_{square}=1.200·3.3=3.96 Ω, $R_{rectangular}$=3.672 Ω. The difference is about 0.3 Ω.

VI. CONCLUSION

The paper describes modeling of textile material from electrical point of view. The structure of fibres is replaced by series-parallel connection. Equivalent electrical circuit is simplified by basic physical laws and resultant resistance is calculated. The model is verified by proposed measurement with the aid of RLCG bridge. Bulk resistivity is measured for two different electrode setting and with and without insulator. Measurement inaccuracy is also discussed. Results show the model is valid, proposed gauging fixture is suitable for measurement and different length of textile sample does not affect resultant bulk resistance.

ACKNOWLEDGMENT

This work was supported by grant no. 2457/2011.

REFERENCES

[1] K. Furuya, Y. Taira, H. Iwasaki, S. Yamamoto, N. Tamaki, T. Harada, and A. Kuramoto, "Wide band wearable antenna for DTV reception," IEEE Antennas and Propagation Society International Symposium, AP-S 2008, pp.1-4, July 2008.

[2] F. Declercq, and H. Rogier, "Active Integrated Wearable Textile Antenna With Optimized Noise Characteristics," Antennas and Propagation, IEEE Transactions on , vol.58, pp.3050-3054, Sept. 2010.

[3] P. Salonen, M. Keskilammi, and L. Sydänheimo, "Antenna design for wearable applications", Tampere university of technology, Finland.

[4] L. Juha, P. Salonen, "On the modeling of conductive textile materials for SoftWearAntennas," IEEE Antennas and Propagation Society International Symposium, APSURSI '09, pp.1-4, June 2009.

[5] N. N. Hoang, J. Miane, and J. Wojkiewicz, "Modeling of Electromagnetic Shielding Effectiveness of Multilayer Conducting Composites in the Microwave Band," Communications and Electronics, ICCE '06, pp.482-485, Oct. 2006.

[6] W.A. Maryniak, T. Uehara, M.A. Noras, "Surface resistivity and surface resistance measurements using a concentric ring probe technique", Trek Application Note Number 1005, 2003, Rev. 1b, pp. 1–4. Available: http://www.trekinc.com/Download/App_note/1005_ Resistivity_Resistance.pdf.

[7] Ubuntu, Package: oregano (0.69.1-1ubuntu1.10.04.1), [online]. 2011 [cit. 2011-03-24]. Available: http://packages.ubuntu.com/lucid-updates/gnome/oregano.

[8] A.J. Perry, B. Inrichen and B. Eliasson, "Fibre diameter measurement by laser diffraction", Journal of Materials Science, vol. 9, issue 8, pp. 1376-1378, 1974.

Telemetry System for Plethysmography and Oximetry Measurement

Marek Penhaker, Vladimir Kasik, Jan Semkovic
VSB- Technical University of Ostrava, FEECS, K450
Ostrava, Czech Republic
marek.penhaker@vsb.cz, vladimir.kasik@vsb.cz, jan.semkovic@vsb.cz

Abstract—**This paper discusses various methods and implementations of telemetry data from the pulse oximeter Criticare 504 USP to computer. The proposal was conditional on sufficient communication speed for the ability to view received data in real time. Part of this work was the design and programming software, allowing the received data to display, save, and then reappear at any time. The result of this work is to create a communication link using low-end part of the cost, both in terms of cost of acquisition, and operational. The aim was to use the most part no longer available. This effort was not enough by the main transmission speed and accuracy of transmitted data.**

Keywords- Telemetry, Plethysmography, Signal;

I. INTRODUCTION

Plethysmography record is a sequence of waves pulse (PV) and a pulse wave is characterized by a cardiac cycle. The pulse wave shape, especially in the systolic part, contributing not only changes in heart rate and amount of blood ejection, but also the aorta and large arteries with elastic properties. The recording waves from the surface of the body Plethysmography method used most often. Learning Plethysmography curve shape are complex and differ from each other both frequencies, the amplitude of both waves and their shape. [1]

Pulse oximeter is a medical device based on microprocessor-based, which measures the percentage of arterial blood oxygen saturation and number of pulses per minute, taking advantage of the principles of spectrophotometric and Plethysmography. The device assesses the transmission of two different wavelengths of light, one of which is in the visible spectrum and the other is in the infrared spectrum. Measurements of light is carried photoemited diode (LED). The light illuminates the hemoglobin in the arterial blood stream and is taken back to the photodiode detector for processing and determination of blood oxygen saturation and pulse values. This investigative method is noninvasive and allows continuous monitoring. [2]

II. II.METHODS

Pulse oximetry company Criticare Systems, Inc.. is available in four different designs. Basic unit includes 504 type pulse oximeter sensor and recording the pulse wave, which is connected a flexible cable with the unit. The widest options available to the device under the name USP-504 contains the pulse oximeter, integrated printer and connection of electrodes for recording single-channel ECG.

The front panel are user interface through buttons, window displays and connecting the two connectors (Fig. 2). One jack is for connecting ECG wires connecting the second sensor and sensing the pulse wave, pulse and oxygen saturation. The rear panel connectors are located on the charging transformer, computer interface RS 232 and two analog outputs.

Figure 1. Front Panel of Criticare 504 USP Plethysmography device [3] https://www.nzma.org.nz/journal/116-1168/297/

The pulse oximeter measures the percentage of arterial oxygen saturation and pulse number using spectrophotometry and Plethysmography. Oxygen saturation sensor contains LEDs that emit two specific wavelengths of light and pulsing through the bloodstream. The sensor is completely noninvasive. There is no source of heat, which could cause a burn patient.

Because O2Hb and Hb absorb light selectively, and some inevitable way, a number of these components can be determined by measuring the intensity of each wavelength that passes through the measuring point. Photodetector, which is located on the opposite side from the light source, measures the intensity of each wavelength that is transmitted through the measuring point. The intensity of light is changed into an electrical signal which enters the oximeter. Effect of skin pigmentation, venous blood and other tissue components are eliminated from the separation of absorption data from pulse field. [3]

A. *Data I/O on medical device*

Pulse Oximeter is equipped with two analog outputs, which can direct signals through the configuration menu. The output signal is continuous through BNC connectors located on the rear panel.

The outputs can be routed all the signals are processed by the device (Table 1). Level output signal is in the range of 0-1 VDC. For precise adjustment of the apparatus senses an analog outputs can use a test signal.

In order to transfer data from a pulse oximeter to a computer, you must first connect the oximeter output to computer input. Pulse Oximeter 504, USP contains two analog and one digital output with RS-232. Computers usually have two RS-232, so the simplest realization of integration would be just over this interface. This link would also have the advantage of simplicity, we received the data sensed Oximeter - plethysmographic curve, saturation and heart rate - over one cable. Unfortunately, we did not receive the ECG. But the biggest shortcoming of this solution is too low quality of the transmitted waveforms, which is probably due to insufficient sampling frequency. [4]

Another option is to use analog output. The signals on these outputs exactly match the measured values, thus avoiding the problem of substandard curves. Computers are not normally have inputs for analog signals and the purchase of converters A / D converter is quite expensive. So it is necessary to find or create the kind of middleman that would transform the analog and digital outputs are sent to the computer input. Possibility of combining the two previous options, or used as an RS-232 and two analog outputs. Plethysmographic curve and ECG would be transmitted through the analog outputs, which would guarantee their accuracy and value of saturation and heart rate via RS-232. Transferring data to a computer would be fully automatic. A practical solution to this problem could be that the analog outputs would be connected to an interface and it would then be further transmitted to the RS-232 computer and RS-232 oximeter is directly connected with the second RS-232 computer. This proposal is quite inefficient, because it is occupied by two RS-232 computer. In some cases, even that solution is impractical because computers normally contain exactly two RS-232 and one of them is sometimes accompanied by a mouse, which is very important to control your computer. Far more effective is to connect analog outputs and RS-232 Oximeter and an interface to transmit data from the relays to the computer via an RS-232. By this increased demands on an interface - a microprocessor, but this seems to be the most effective solution of all.

III. DESIGN AND SOLUTION PROPOSAL

The link between pulse oximeter and the computer is realized by the PIC16F873 microprocessor. For its function must be properly connected and have attached themselves to supporting districts to ensure its proper operation. All these components are integrated into a single printed circuit, which is set in a closed box with five pins:

- Analog input plethysmographic curve – BNC
- Analog input for ECG curve – BNC
- input for receiving data from the RS-232 Pulse Oximeter - Cannon 9 type connector Female
- output for sending data to the RS-232 PC - Cannon 9 type connector Female
- power supply

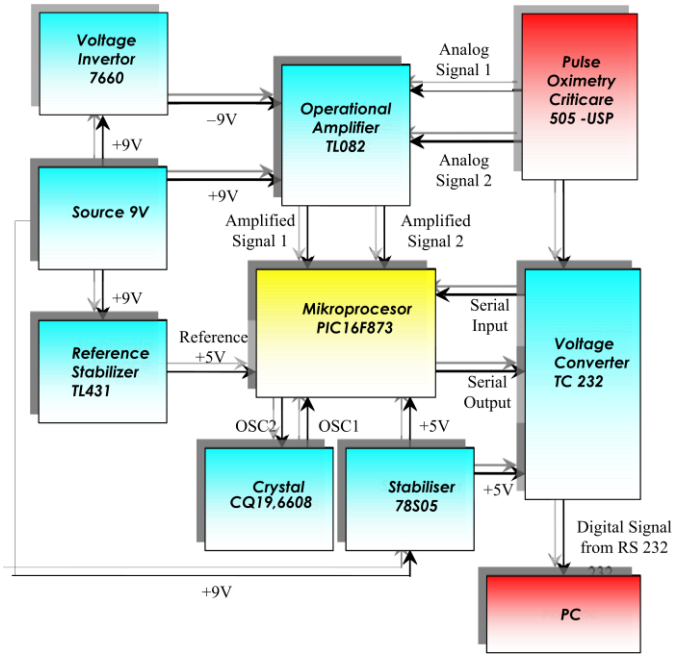


Figure 2. Diagram of Interconnection PC and 504 USP Plethysmograph

A. *Power Supplying*

For power supply circuits are needed between 4.5 and 5.5 V. For this purpose, the source used to transfer the line voltage (220V) to 9V (resource that would transform the mains voltage directly to the desired voltage between 4.5 and 5.5 is not widely available). As soon as this source is connected 78S05 stabilizer, which stabilizes any input voltage between 6.5 to 30V (12V in our case) to 5V. The maximum current consumption of the stabilizer is 1.5 A, which is more than enough. [5], [6]

B. *Reference Voltage Source*

PIC16F873 microcontroller has the ability to connect the maximum and minimum reference voltage for A / D converter. Given that output analog signals from the pulse oximeter range from 0 to 1V, the connection of the minimum reference voltage nil (in this case is regarded as the minimum ground or 0V), while the connectivity of the maximum reference voltage contrast is used (in To do otherwise would be considered a maximum reference voltage power supply, or 5V, which would range A / D converter largely untapped). The most ideal solution would be to connect a 1V maximum reference voltage. Unfortunately, as specified by the A / D converter of the microprocessor is not possible to connect the reference voltage is less than the supply voltage minus 2.5 V, or the maximum

reference can not be less than 2.5 V. For this reason, it was necessary to first amplify the input signal and then routed to the input. Gain was set at 5, or the input voltage between 0 - 1V to step up the voltage 0 - 5V. In this case, the theory may not be specifically a maximum reference voltage being supplied, as supply voltage is 5V. However, stabilizer 78S05 sufficient accuracy is not guaranteed, so the stabilized voltage may not be exactly 5V (measured it was found that the connected stabilizer 78S05 stabilizes the voltage 4.93 V at any allowable current consumption).

As a reference voltage source was used TL431 reference voltage stabilizer. This stabilizer is connected to a potentiometer, which allows you to accurately set the reference voltage in the range from 2.5 to 36V. Involvement of the reference voltage stabilizer TL431 is as follows:

C. Operational Amplifier

For the amplification of both input signals from 0 - 1V to 0 - 5V circuit TL082 was used. This circuit includes two operational amplifiers (AMP A and AMP B), which are based on technology, BI-FET II. They are also suitable for use in applications where the input signal is changing rapidly. The advantage is the possibility of exchanging this amplifier for more accurate, because a lot of operational amplifiers is designed as pins, which are as follows. [7]

Both operational amplifiers are connected as a noninverting amplifier with amplification 5 (Fig. 8). The problem is supply, which must be balanced with operational amplifiers, ie. on a single supply input voltage must be positive (eg 9V), the latter must be the same, but the negative voltage (-9V). There are amplifiers with unbalanced power, but they are less accurate, especially when connected to the input voltage of zero, which is still on the low voltage output. The circuit was therefore necessary to create a negative voltage power operational amplifier.

D. Negative Violtage Realising

To create a negative voltage power operational amplifier circuit is used 7660th This circuit when the voltage between 0 - 10V input to create its output voltage to exactly the opposite (or 0 to-10V). [8]

E. Oscilator Inductink

To operate the microprocessor internal oscillator must be accompanied by a parallel crystal or resonator. Crystal connection between the pins OSC1 and OSC2 microprocessor. In our case it is used in parallel CQ 19.6608 crystal, whose frequency is 19.6608 MHz.

F. Complete Circuit design

The block diagram of the link between pulse oximeter, Criticare 504 USP and a computer. All these signals are related to a common ground. Overall diagram contains the elements shown in Figure No. 3

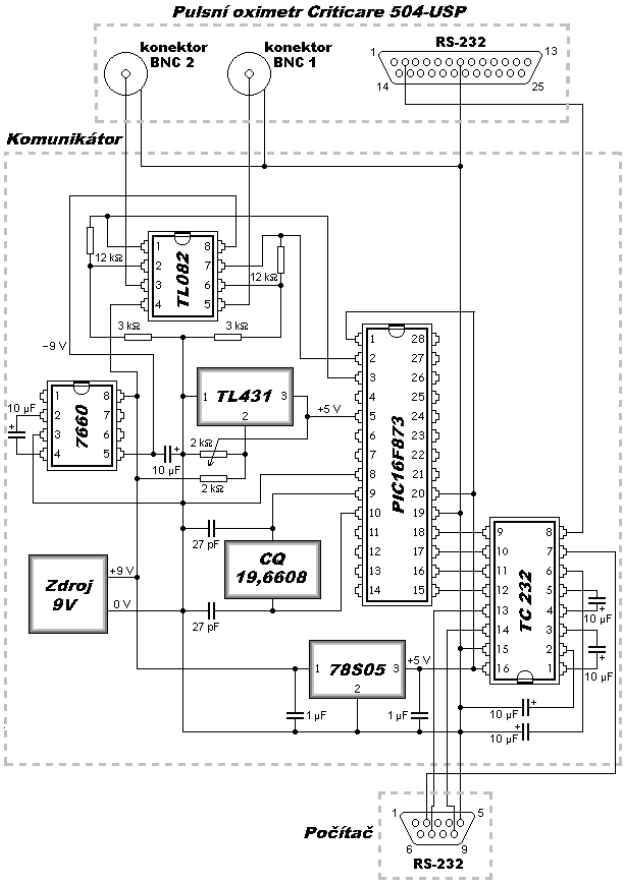


Figure 3. Diagram Complete Electronics Circuits In Telemetry Communicator

IV. PLETHYSMOGRAPY RECORD DATA MEASUREMERNT

To receive data sent from the pulse oximeter through a PIC16F873 microcontroller to a computer was necessary to create a program that they would receive and process the data. This program must meet certain criteria. It should be as simple and easy to understand and control. Furthermore, it should appear accepted Plethysmography curve and ECG in real time. But the main requirement was the ability to store measured data, and thus also creating a database of patients and their individual tests and measurements.

The entire program was programmed in a programming environment Borland Delphi version 5 This programming environment has a great support not only databases, but also a wide range of standard and nonstandard controls typical of all Windows versions.

V. DISCUSSION

Proposals to link the data-puslní oximeter to a computer, it is certainly many. One option would be to purchase a new pulse oximeters, which strongly supported directly communicate with the computer. But do not except the possibility of new oximeters far such enhancements to its functionality to new devices due to cost-effective. Far cheaper and more effective to create a custom link Oximeter already purchased a computer. Further consideration was given to

minimize the costs associated with the acquisition of the necessary components for implementation of the connection. But the most important factor was the accuracy and speed and transmission quality.

For processing data transmitted from the pulse oximeter to a computer program was developed in the Delphi programming environment version 5 The program is designed to be easy to use and the actual measurement of the most automated. The operator does not need to pay close attention to its control and may be given to the patient and measurement. The program is able to automatically find the serial port it is connected to the communicator, the ongoing investigations will automatically switch between the measurements to be within the specific tests performed. Saved curve can be viewed at any time, you can have your view and any of them to move cursors, for example, you can use to determine the time difference between the two peaks in the graph, the curve can be printed. To store the measured data were used in the Paradox database, type 7, which have become standard in the database. This allows the use of these data in other programs, or transfer of such databases to the databases of other types, in future it will be possible to program more effective programs and use those already measured and stored data. In the case of the adoption of clear criteria for assessing plethysmography curves could be the new programs enrich the analysis, thereby increasing the value of information stored data. It is also possible that in future this data will become part of a wider database includes not only data plethysmography testing, but perhaps also as data from ECG, EEG and, if not all examinations, the patient undergoes. This would allow doctors to keep track of the overall health of the patient (and not just over the state relating to physician specialty) and more efficiently, accurately and quickly determine the diagnosis.

VI. CONCLUSION

The goal of this work was to create a computer as a terminal for the pulse oximeter 504 USP-fy. Criticare. The pulse oximeter is designed for standalone operation without the support of a computer, but your work computer with this device and physicians significantly simplifies and accelerates work with measured data. Not to say that the use of computers has brought entirely new possibilities in Plethysmography. This is mainly due to the fact that the plethysmographic curve assesses each doctor only visually without clearly identified the analysis. Until it adopted a unique apparatus for assessing pletysmografických curves, then in this area will only serve as a computer terminal and storage space for the measured data without the possibility of increasing the contribution of information value of the measured curves using different analysis. Yet it is unquestionably computer irreplaceable and important (and today have been directly necessary) relief work. It brings in addition to specific practical applications directly in the hospital are also indications of possible future direction of the link between medical devices and computers. There was a link to only one device with a computer. But the speed of computers, microprocessors and communications continues to grow. In the future could be linked together several devices at once with one computer, either through another intermediary, or even if it still leaves the computer and device support.

Acknowledgment

The work and the contribution were supported by the project: Ministry of Education of the Czech Republic under Project 1M0567 “Centre of applied Cybernetics”, student grant agency SV 4501141 “Biomedical engineering systems VII” and TACR TA01010632 “SCADA system for control and measurement of process in real time”. Also supported by project MSM6198910027 Consuming Computer Simulation and Optimization.

REFERENCES

[1] Korpas, D., Halek, J. Pulse wave variability within two short-term measurements. (2006) Biomedical papers of the Medical Faculty of the University Palacký, Olomouc, Czechoslovakia, 150 (2), pp. 339-344 ISSN: 12138118

[2] Davies, G. Gibson, A.M., Swanney,M., Murray,D., Beckert, L., Understanding of pulse oximetry among hospital staff Journal of the New Zealand Medical Association, 24-January-2003, Vol 116 No 1168

[3] Krejcar, O., “Problem Solving of Low Data Throughput on Mobile Devices by Artefacts Prebuffering”, In EURASIP Journal on Wireless Communications and Networking, Article ID 802523, 8 pages. Hindawi publishing corp., New York, USA, DOI 10.1155/2009/802523 (2009)

[4] Skapa, J., Siska, P., Vasinek, V., Vanda, J. Identification of external quantities using redistribution of optical power - art. no. 70031R In Book OPTICAL SENSORS 2008 Volume 7003, p. R31-R31, 2008,ISSN: 0277-786X , ISBN: 978-0-8194-7201-4

[5] Simsik, D., Galajdova, A., Majernik, J., Hrabinska, I., Zelinsky, P.: The video analysis utilization in rehabilitation for moblity development. Lékař a technika. Česká republika, 4-5, Ročník 35, 2004, ISSN 0301-5491, pp. 87 - 92.

[6] Zivcak, J., Petrik, M., Hudak, R., Toth, T., Knezo, D., Kovalova, E. Embedded tensile strenght test machine FM1000 - An upgrade of measurement and control, Diffusion and Defect Data Pt.B: Solid State Phenomena 147-149, pp. 657-662, ISSN: 10120394, ISBN: 3908451655; 978-390845165-

[7] Klima, M., Pazderak, J., Bernas, M., Pata, P., Hozman, J., Roubik, K. Objective and subjective image quality evaluation for security technology, IEEE Annual International Carnahan Conference on Security Technology, Proceedings , pp. 108-114

[8] Garani, G. Nest and unnest operators in nested relationsData Science Journal 7, pp. 57-64 ISSN: 16831470, DOI: 10.2481/dsj.7.57

[9] Aidu, E.A.I., Trunov, V.G., Titomir, L.I., Tysler, M., Turzova, M., Szathmary, V. Electrocardiographic ST segment changes as an indicator for localization of injury potentials. A computer simulation studyKardiologia 15 (1), pp. 21-24 ISSN: 12100048

[10] Havlík, J., Uhlir, J., Horcik, Z. Thumb motion classification using discrimination functions International Conference on Applied Electronics 2006, AE , art. no. 4382963, pp. 55-57, ISBN: 8070434422;978-807043442-0, DOI: 10.1109/AE.2006.4382963

[11] Hozman, J., Zanchi, V., Cerny, R., Marsalek, P., Szabo, Z. Precise Advanced Head Posture Measurement In Book Challenges in Remote Sensing - Proceedings of the 3rd Wseas International Conference on Remote Sensing (REMOTE '07) p. 18-26, 2007, ISSN: 1790-5117, ISBN: 978-960-6766-17-6

[12] Cerny, M. Movement Activity Monitoringof Elederly People – Application in Remote Home Care Systems In Proceedings of 2010 Second International Conference on Computer Engineering and Applications ICCEA 2010, 19. – 21. March 2010, Bali Island, Indonesia, Volume 2NJ. IEEE Conference Publishing Services, 2010 p. ISBN 978-0-7695-3982-9

Verify the Accuracy and Time Stability of the Radiation Labels

Zuzana Vasickova [1], Marek Penhaker [1], Jan Skapa [2]

[1] Department of Measurement and Control, VSB-Technical University of Ostrava, Ostrava, Czech Republic
[2] Department of Telecommunication, VSB-Technical University of Ostrava, Ostrava, Czech Republic

Abstract— **This paper deals with the measurement of performance and stability over time of irradiation used to label the blood preserved as an indication of killing white blood cells. The retention of blood products causing autoimmune response of white blood cells self-destruct and, if so the white blood cells were removed. Currently on the market two types of irradiation labels to indicate these changes proportionally. The work deals with the measurement reliability and temporal stability of these indicators.**

Keywords— **Indicator, Blood Cells, Transfusion**

I. Introduction

Collected blood is taken into plastic sacs where it is stored until its application to the recipient. The donor´s blood includes white blood cells – leukocytes, which usually participate in immune system functioning. It means that the white blood cells are usually able to fight viruses, bacteria and other pathogens or elements as well as tumor cells and strange substances for the organism at all. That phenomenon may cause undesirable reactions after the transfusion and the immunization. It originates from the transfusion of the viable T-lymphocytes proceeding from the donor to the immunologically weakened recipients. Transfunded lymphocytes proliferate (reproduce to excess) and react against the recipient because his defective immunity isn´t able to react against the extraneous lymphocytes and to eliminate them. As a precaution of these undesirable reactions the blood products are irradiated. The irradiation effect is in killing T-lymphocytes whereas thrombocytes´ viability is preserved. The radiation dose for the transfusion preparation shouldn´t drop below 25 Gray and reach over 50 Gray. The exposition length of the transfusion preparation depends on the type and power of the source (depending up the source aging the length of irradiation increases). Indications for the blood products irradiation are: immune deficiency, high-dosed chemotherapy, organ transplantation, intrauterine and infantile (postneonatal) transfusions, Hodgkin´s disease, transfusions by recipient´s blood relatives and the whole body irradiation.

A. The indicators

Indicators are used for control of appropriate blood products radiation dose. The indicator is stuck on the blood bag before its irradiation and it stays there until the blood product is used. There are two types of the indicators available in the Czech Republic: Rad Sure a Rad Tag.

Transfusion-associated graft-versus-host disease (TA-GVHD), a rare complication of blood transfusion, is usually fatal. Patients in certain clinical categories, including bone marrow transplants, solid tumors and acquired T cell defects, are at greater risk of developing TA-GVHD. TA-GVHD can be prevented by the irradiation of cellular blood products prior to transfusion. The RAD-SURE blood irradiation indicators provide positive, visual verification of irradiation. When attached to blood products, RAD-SURE indicators show whether the blood products have been irradiated or not.

Before a blood product and its attached indicator are irradiated, the indicator reads, "NOT IRRADIATED." After the blood product and its attached indicator are irradiated, the word "NOT" in the indicator window is obscured and the indicator reads, "IRRADIATED." RAD-SURE indicators only indicate that irradiation has occurred. They should not be used as dosimeters to measure the dose delivered by the irradiator. RAD-SURE Type 15 Gy and Type 25 Gy indicators should only be used with irradiators having cesium-137 or cobalt-60 radiation sources, or other sources producing radiation of equal or greater energy. There are examples of the indicators before and after the irradiation in the Figure 1 and Figure 2. [1]

Fig. 1 The Rad-Sure Indicator, not irradiated

Fig. 2 The Rad-Sure Indicator, irradiated (35Gy)

RadTag, an innovative blood irradiation indicator gives more than just a yes or no. It indicates if the product received the minimum dose (25Gy), or exceeded the maximum allowable dose of 50Gy. Simply confirm the sensitive blue dot is the same as or darker than the 25 Gy colour and if necessary, not darker than the 50 Gy colour. The RadTag Irradiation Indicator is attached to a blood bag prior to the bag's placement in an irradiator. The sensitive portion of the indicator is white initially and changes to a blue colour upon exposure to ionizing radiation; the shade of blue is an approximation of the amount of radiation dose delivered. The colour change is permanent and immediate and does not require any further processing or development. The amount of colour change has been optimized to give a clear indication that the blood bag has been processed at levels of irradiation currently used in the industry. A typical processing unit will use 35 Gy as a specified dose delivered to the central portion of the irradiation container, and 25 Gy is the minimum dose at any other point.

[2]

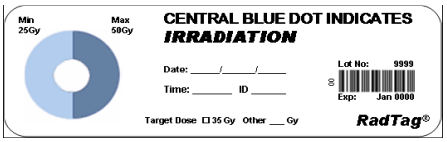


Fig. 3 The Rad-Tag Indicator, not irradiated

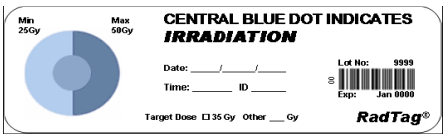


Fig. 4 The Rad-Tag Indicator, irradiated (35Gy)

II. Methods

The aim of the study was to verify the utility of the indicators, their stability after the irradiation, influence of light sources and mutual comparison of two types of indicators.

A. The Single-shot test

This type of test was used to mutual comparison of two types of indicators. The both indicators Rad sure and Rad tag were used in common irradiation cycle of blood bags. The photo documentation of indicators was made before and after the irradiation cycle. And we compared the evaluation of the hospital stuff.

B. The long-term test

These tests were used to study the stability of indicators while the indicators are exposed to the radiation of light. The indicators are exposed to normal radiation of light before and after the irradiation in common conditions. If the indicators change their color due to the incident lighting it may lead in wrong results.

The blood bags are stored in fridge with stable lighting after the irradiation and before the transfusion. If the indicators darken during this storage it may lead in wrong interpretation. The stuff could evaluate this like oversized irradiation. The distanc between indicators and source of light was 10-15cm.

III. Tests and results

A. The Single-shot test

The blood products are irradiated by dose 32-35 Gy there in the blood center. That dose is achieved after 7 min. 50 sec. irradiation length. We wanted to try whether the indicators wouldn´t indicate the irradiation even when the absorbed dose would be lower. Therefore we halved the irradiation time (3 min. 55 sec.). The indicators wrongly showed irradiation over 25 Gy! In doing so the absorbed dose should be 16-18 Gy. The staff that evaluates the indicators commonly (nurses) marked both indicators as fully irradiated. The work procedure was kept in full, the setting was made by fully trained and experienced staff. The indicators were in good order before the irradiation and they were kept cold. The comparison of 16Gy and 32Gy irradiation: the RD indicator showed the same outcome in either event. It means the totally black coloring. The RT indicator darkened more in 32Gy than in 16Gy.

We supposed that the indicators may absorb any irradiation dose during the check at the airport (X-ray testing). We repeated the tests with indicators which were delivered in leaden box. The results were better. The Rad Sure indicators indicated the irradiation dose exactly. The Rad Tag indicators still indicate more absorbed irradiation, but less then in the first case.

Fig. 5 The indicator Rad Tag, irradiated 32Gray, first test

Fig. 6 The indicator Rad Tag, irradiated 16Gray, first test

Fig. 7 The indicator Rad Sure, irradiated 16Gray, first test

Fig. 8 The indicator Rad Tag, irradiated 16Gray, second test

Fig. 9 The indicator Rad Tag, irradiated 25Gray, second test

B. The long-term test A

The not-irradiated indicators were stored in a fridge with stable lighting for 30 days. 1RS and 1RT by their front side to the light source. 1RS and 1RT by their back side to the light source. 1RS and 1RT stored in a dark wrapper due to the reference. They were checked every day; the scans were made once a week.

The RadTag indicator, which was stored for 30 days below a light source, darkened softly. The color change was similar to color change of indicator after irradiation dose 6Gray. That colour change can lead to wrong evaluation of the indicator. The indicators Rade Sure had the still color during the whole test.

Fig. 10 The indicator Rad Tag, not irradiated, after 30days below the lighting.

C. The long-shot test B

2RS and 2RT indicators were put at each other so that they absorbed the same dose, and they were put into the blood products irradiator in a common blood product irradiation cycle. The set absorbed dose was 32-35 Gy.

Consequnetly the irradiated indicators were used for the long-term test of stability:

Irradiated 1RS and 1RT, non-irradiated 1RS and 1RT were put in a fridge by their front side to the light source.

Irradiated 1RS and 1RT, non-irradiated 1RS and 1RT stored in a dark wrapper due to the reference.

They were checked every day, the scans were made once a week.

The indicators Rade Sure had the still color during the whole test. The RadTag indicator, which was stored for 30 days below a light source by their front side, darkened softly.

Fig. 11 The indicator Rad Tag, irradiated 32Gray

Fig. 12 The indicator Rad Tag, irradiated 32Gray, 35th day below the lighting.

IV. CONCLUSIONS

The blood irradiation indicators are used for control of appropriate blood products radiation dose. The indicator is stuck on the blood bag before its irradiation and it stays there until the blood product is used. The aim of this work was to verify the utility of the indicators, their stability after the irradiation, influence of light sources and mutual comparison of two types of indicators. The results of performed tests are: The incidence of lighting affected the Rad Tag Indicators. The RadTag indicator, which was stored for 30 days below a light source, darkened softly. The color change was similar to color change of indicator after irradiation dose 6Gray. The indicators Rad Sure had the still color during the whole tests. Another result is that the transport of indicators may also influence them. Probably during the check in at the airport (X-ray testing) the indicators absorb irradiation which influences another using of them.

ACKNOWLEDGMENT

Grand – aided student, Municipality of Ostrava, Czech Republic. The work and the contribution were supported by the project Grant Agency of Czech Republic – GAČR 102/08/1429 "Safety and security of networked embedded

system applications". Also supported by the Ministry of Education of the Czech Republic under Project 1M0567.

References

1. Blood irradiation indicator Rad Sure- Brochure at http://www.lmb.de
2. Product Information Sheet Rad Tag at http://www.radtagtech.com

Author: Zuzana Vasickova, Msc
Institute: Department of measurement and control, VSB-TUO
Street: 17. listopadu
City: Ostrava
Country: Czech republic
Email: zuzana.vasickova@vsb.cz

ECG Signal Simulator with Programmable Hardware

Vladimir Kasik[1]

[1]VSB – Technical University of Ostrava, FEECS, DMC
Ostrava, Czech Republic
vladimir.kasik@vsb.cz

Abstract **– This paper is aimed to Electrocardiogram (ECG) signal simulator design for biomedical applications. The design is based on the low-power Complex Programmable Logic Device (CPLD) for easy functional modifications and long battery operation. The solution is an alternative to a simulator using microcontroller architectures. The advantage is high flexibility of hardware solutions and thus easy possibility of further technical adjustments directly based on the philosophy of CPLD.**

Keywords- Biological signal; CPLD; ECG monitor; Simulator

I. INTRODUCTION

The article introduces the biological signal simulator, which in conjunction with any ECG monitor enables visual interpretation of ECG waveforms.

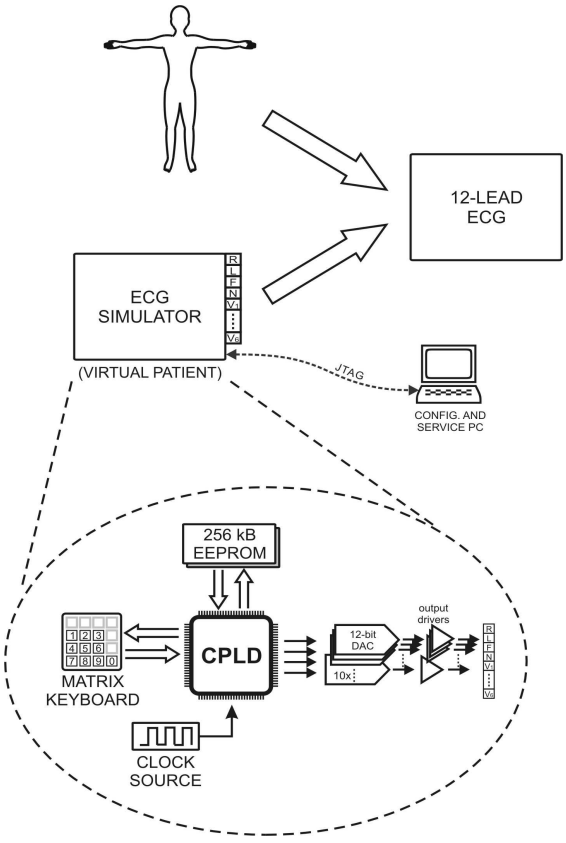


Figure 1. Block diagram of the biological signal simulator unit

In the following text a structure of the biological signal simulator is described with emphasis on Programmable Logic Design. Particular design components and also design technique of the project are chosen with respect to low power consumption.

All the design concept of the unit is aimed to low power consumption, reliability, safety and appropriate functionality. In contrast to classical solution with analog circuits or microcontroller, a CPLD [1] is used. The block diagram of the simulator unit is shown in Figure 1.

II. ECG SIMULATOR ARCHITECTURE

As shown in Figure 1, the unit consists of these parts:

CPLD device – the main logical part of the unit, which establishes an interconnection with other blocks of the circuitry. It includes digital design as shown in Figure 1 as combinatorial and sequential logic.

The row and column lines are designated for 4x4 matrix keyboard, however only 10 keys are currently used for the ECG waveform selection.

EEPROM – contains the waveform samples for D/A converters. The minimal memory capacity is 256 kB with store organization 256k x 8. However it can be as large as the CPLD logic resources make it possible, for example, to units of megabits [2]. The EEPROM size is then limited by the customers requests concerning to the maximum time – length of the waveforms and the number of artefacts. The parallel type EEPROM can have 8- or 16-bit data bus and the supply voltage is limited to 3,3V (Low Voltage). Due to slow sample rate of the ECG signal simulator the data throughput of the EEPROM is not a critical feature. The control and address signals for the memory are periodically generated in the memory controller inside the CPLD.

D/A converters – The simulator is equipped with 10 independent analog outputs, therefore it uses two 12-bit octal DACs [3].

Output drivers – They adjust output impedance and voltage range to meet the ECG specification.

Clock source – All the digital design is clocked from single quartz oscillator. The frequency 1.843 MHz is divided in CPLD to several frequencies needed for clocking DACs , keyboard and memory access.

Matrix keyboard – the number of buttons corresponds with the possible artefacts of the simulator. Simple pressing the button changes the output waveform immediately. Inside the CPLD device the matrix keyboard is connected to a keyboard scanner, which produces periodic testing signals after a key is pressed, as shown in Figure 2.

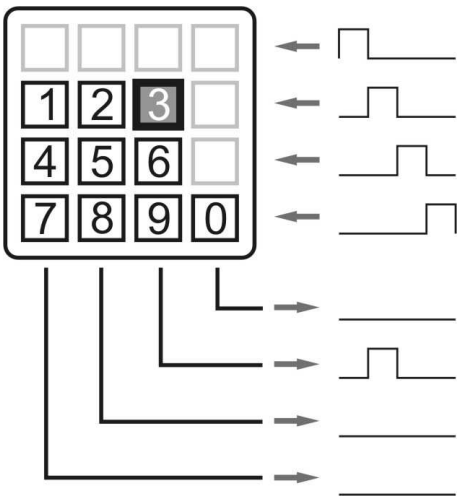


Figure 2. Keyboard control – key '3' pressed.

Before any keypress the row lines are all in a high logic level due to minimizing the dynamic power consumption.

Main part of digital design is implemented in Xilinx CoolRunner II CPLD. That device family is chosen primarily due to low power consumption, which depends on the clock frequency and designed logic structure. The design entry is hierarchical and is described in VHDL language. The digital circuitry implemented in programmable logic contains three concurrent modules: Keyboard scanner, Clock divider and DAC + Memory controller.

Keyboard scanner scans the button pressed and encodes that information into four-bits binary index. That module works sequentially and it is able to scan up to 4x4 matrix keyboard, however the simulator unit needs only ten buttons (3x4 keyboard).

Clock divider divides the clock signal from external input to frequency needed for keyboard scanning process and especially for DAC + MEMORY Controller.

DAC and memory controller periodically generates addresses for the memory, receives its data (waveform samples) and generates appropriate signals for D/A converters. Address range is determined by the information from the keyboard scanner.

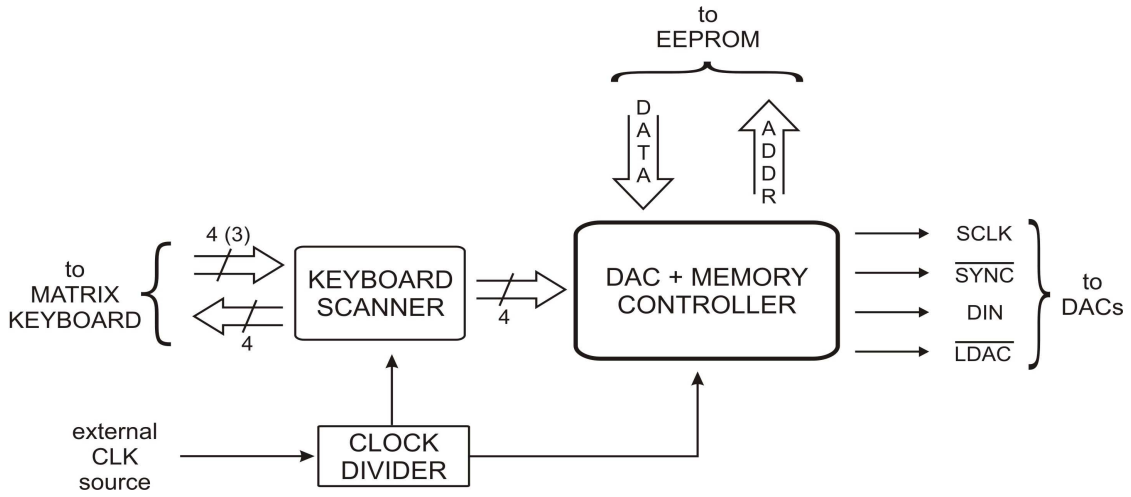


Figure 3. Digital design in CPLD

In addition, the last sample in that range is identified with a special word 0xFFFF in the memory. D/A converters used in that design use an SPI interface and therefore the DAC controller generates SCLK, SYNC, DIN and LDAC signals.

III. Discussion

There is one part in the design, not presented in previous text, which has two alternatives in this time – an indication of the selected waveform. Suggested alternatives are low power LED diodes and simple LCD display without controller. Advantage of the LEDs is the visibility in the dark and CPLD design simplicity (simple discrete outputs for LEDs), while simple LCD has the feature of negligible supply current and less output pins needed. The resulting Design Utilization Summary in Table 1 relates to simple LED outputs, however in this time the LCD output feature is also designed.

TABLE I. Design Utilization Summary for XC2C256-6-VQ100 CPLD

Feature	CCD Sensor	CMOS Sensor
Macrocells Used	242 / 256	95 %
Pterms Used	266 / 896	30 %
Registers Used	71 / 256	28 %
Pins Used	76/80	95 %

IV. Design Verification

The design has been verified on Xilinx's CPLD II Development Kit connected to experimental DAC board with amplifiers and other analog circuitry. Samples data were obtained either from any sample databases of the hospital or by digitizing of specialized scientific resources (books).

The signal period can be easily adjusted by the clock frequency at all waveforms (CLK division ratio). Some problems have been identified in the analog part of the simulator with the amplifiers. Their in-circuit feedback needs to be precisely calibrated for appropriate multi-channel output. The Figures 4-7 shows sample data of several artifacts. The vertical axis of graphs represent the numerical value of samples, the index is on the horizontal axis.

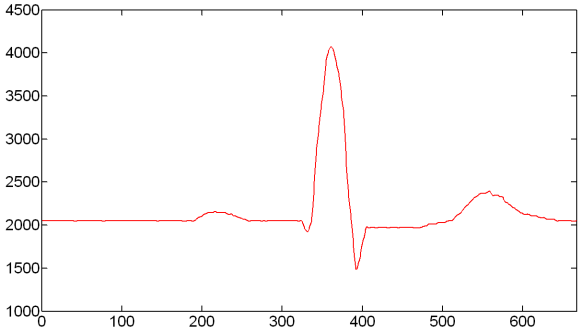


Figure 4. Sample of generated signal – normal ECG waveform

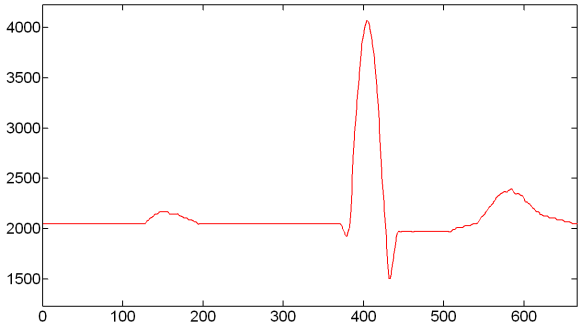


Figure 5. Sample of generated signal – variation on P-wave

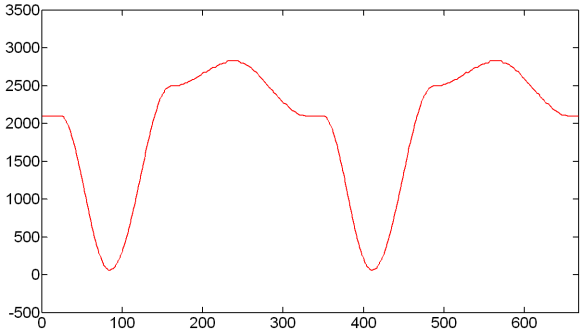


Figure 6. Sample of generated signal – ventricular tachycardia

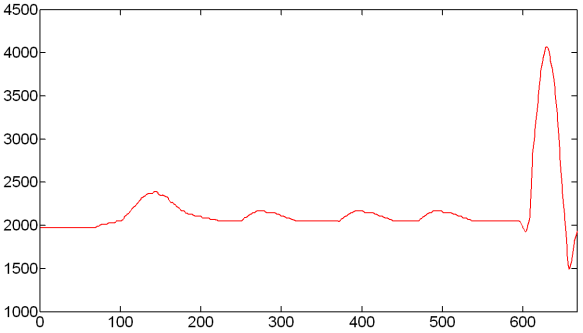


Figure 7. Sample of generated signal – atrial fibrillation

V. Conclusion

The project introduces new innovative approach to biological signal simulator design with low power CPLD device. The main features of that simulator are 10 analog outputs, 10 artefacts and up to 30 seconds of time period. However these parameters depend on the EEPROM / FLASH memory used. The capacity can achieve up to tens of megabytes. Reprogrammability of the CPLD and EEPROM enables to update the major part of design in the future. Aditionally, there is the possibility to generate 1mV calibration signal, which can be used to adjust sensitivity of the sensor, thus the ECG monitor and recorder attached to 1mV standard of 10 mm deviation of recording tracks. An SMT technology is used for the layout, what helps to realize the ECG simulator in the hand-held size. It can be utilized in many applications from hospitals

through test services up to applications as learning tool in a health services [4], [5], [6], [7].

ACKNOWLEDGMENT

The work and the contribution were supported by the project: Ministry of Education of the Czech Republic under Project 1M0567 “Centre of applied cybernetics”, student grant agency SV 4501141 “Biomedical engineering systems VII” and TACR TA01010632 “SCADA system for control and measurement of process in real time”. Also supported by project MSM6198910027 Consuming Computer Simulation and Optimization.

REFERENCES

[1] Xilinx, Inc: CoolRunner-II CPLD Family [online]. [cit. 2009-05-06]. URL: <http://www.xilinx.com/support/documentation/data_sheets/ds090.pdf>.

[2] AT49BV322DT-70TU — ATMEL — FLASH MEMORY [online]. [cit. 2010-03-20. URL: <http://uk.farnell.com/1455055/semiconductors-integrated-circuits/product.us0?sku=atmel-at49bv322dt-70tu>.

[3] AD5308 2.5 V to 5.5 V Octal Voltage Output 8-Bit DACs in 16-Lead TSSOP [online]. [cit. 2008-05-20]. URL: <http://www.analog.com/en/prod/0,2877,AD5308,00.html>

[4] M. Stankus, M. Penhaker, J. Kijonka, P. Grygarek, “Design and Application of Mobile Embedded Systems for Home Care Applications,” In Proceedings of 2010 Second International Conference on Computer Engineering and Applications. ICCEA 2010, 19. – 21. March 2010, Bali Island, Indonesia, Volume 1NJ. IEEE Conference Publishing Services, 2010 p. 412-416. ISBN 978-0-7695-3982-9, DOI: 10.1109/ICCEA.2010.86

[5] O. Krejcar, D. Janckulik, L. Motalova, K. Musil, M. Penhaker, “Real Time ECG Measurement and Visualization on Mobile Monitoring Stations of Biotelemetric System,” In The 2nd Asian Conference on Intelligent Information and Database Systems, ACIIDS 2010, March 24-26, 2010 Hue City, Vietnam. Springer, Advances in Intelligent Information and Database Systems, Studies in Computational Intelligence, SCI Vol. 283. pp. 67-78. N.T. Nguyen et al. (Eds.). Springer-Verlag, Berlin, Heidelberg. ISBN 978-3-642-12089-3, ISSN 1860-949X (Print) 1860-9503 [Online], DOI 10.1007/978-3-642-12090-9_6

[6] Z. Labza, M. Penhaker, M. Augustynek, D. Korpas, “Verification of Set Up Dual-Chamber Pacemaker Electrical Parameters,” In 2010 The 2nd International Conference on Telecom Technology and Applications, ICTTA 2010. March 19-21, 2010, Bali Island, Indonesia, Volume 2NJ. IEEE Conference Publishing Services 2010, p. 168–172. ISBN 978-0-7695-3982-9., DOI: 10.1109/ICCEA.2010.187

[7] M. Penhaker, M. Rosulek, M. Cerny, L. Martinák, “Design and Implementation of Textile Sensors for Biotelemetry Applications,” In IFBME Proceedings of the 14th Nordic – Baltic Conference on Biomedical Engineering and Medical Physics, 16.-20.6.2008 Riga, Latvia. Berlin Heilderberg New York: Springer et IFMBE, 2008. s.54 et 405-408 volume 20. ISBN 978-3-540-69366-6, ISSN 1680-0737.

Digital Control of RPLIS2048 Line-Scan Sensor using FPGA

Vladimir Kasik[1], Ivan Stancl[2]

[1,2]Department of Measurement and Control
VSB – Technical University of Ostrava, FEECS, DMC
Ostrava, Czech Republic
vladimir.kasik@vsb.cz

Abstract – **This paper describes the design of digital control line-scan image sensor RPLIS2048. To ensure sufficient speed for control and image processing an FPGA device was used. The document describes an internal structure of the digital logic and communication between the line-scan sensor and FPGA. Verification of sensor control functions is demonstrated on a test pattern. To display the processed image an FPGA development kit with VGA output is used.**

Keywords-FPGA; CMOS; Image Sensor; Line Scan Camera; Nexys2

I. INTRODUCTION

The project described in this paper builds in part on experiments with a line image sensor RPLIS2048 by the author, in which it was designed a prototype of line camera module with that sensor and the LPC2138 ARM7 microcontroller series. This module was used to verify basic functions and design of the sensor management algorithm for the microcontroller.

Although experiments have confirmed the functionality of the sensor module and principled application of the proposed construction, they have also shown some disadvantages of microcontroller control of the sensors in the proposed structure. Generating the clock signal by the microcontroller has proven problematic, since the maximum frequency of reading the image had to be a relatively large proportion of computer time devoted to the microcontroller generating a trivial task fundamental waveform signals, the tuning step program when it was interrupted by generating the clock signal and image sensor and getting beyond the recommended working conditions.

As a not very satisfactory options were evaluated microcontroller applications in terms of generating precisely timed and sufficiently fast control signals, while the role of utilization of real-time transmission of image data and also in other more exacting tasks.

These results led to believe that control of the sensor is more appropriate with use of programmable logic CPLD in the simpler cases, and FPGA for larger and more complex applications.

FPGA based sensor control brings computing speed, precise synchronization of multiple sensors and the parallel processing of data. Using inexpensive CMOS sensor enables use of the system in the measuring position, robotics, industry and other fields [1], [2], [3].

In this project, the existing prototype module with an image sensor is used (after necessary improvements) as the basis for a new camera module designed for connecting to a development board with Xilinx FPGA.

II. APPLICATIONS OF THE LINE IMAGE SENSORS

Line-scan image sensors are used in many industrial, commercial and scientific applications.

In the industry the line cameras are used for optical inspection of products on production lines, reading bar codes and check prints. Special application of the line cameras is capture a panoramic image by rotating the camera [4], [5]. Very widespread use of image sensors is line scan paper applications, like the scanners, fax machines and automatic document readers. In similar applications, there is often used a module for contact scanning of the documents containing not only the sensor, but also LED illumination and built-in optical system. Also, there are often used so-called gradient index microlenses..

Another example of the line sensor applications is a position encoding using code template (code-disc or belt regarding the type of encoder), usually with a Gray code.

Figure 1 shows the application of line sensor for measuring distances on the principle of triangulation.

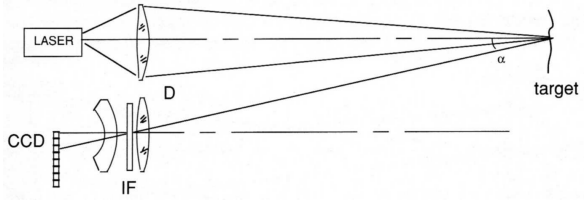


Figure 1. Triangulation-based distance measurement using line-scan image sensor and a laser [6].

Range of other applications includes use in spectroscopy, in robotics or for non-contact measurement of dimensions and location. Examples of such applications can be found in [7], [8], and in many other resources.

III. LINE-SCAN CMOS SENSORS AND THEIR CONTROL

Line-scan CMOS sensors are optoelectronic components that convert incident radiation in the visible or infrared spectrum to a series of voltage values corresponding to the energy incident to individual pixels. Unlike the more common area image sensors they provide only one-dimensional visual information - the sensitive area is formed by a single line of pixels. This allows a complete reading of image data and processing in a very short time compared with the image sensor circuit, which is advantageous for many applications. Current sensors achieve a maximum frame rate up to hundreds of thousands of lines per second.

The typical internal structure of the sensor and control signals is shown in Figure 2. CLK signal is used to read data clocking, the START signal starts reading the image data and usually determines the start and end of the integration interval of the sensor (exposure time). Reading image data are available on the VIDEO OUT. END OF SCAN signal indicates the completion of scanning the entire image.

Besides the usual signals of certain sensors there are other signals used for the programming and configuration of sensors, switch into standby mode, etc. The basic arrangement of the sensor is illustrated with the analog voltage output, which needs to be A/D converted for next digital processing. Some sensors include the A/D converter on the chip. Besides the usual signals of certain sensors there are other signals used for the programming and configuration of sensors, switch into standby mode, etc. The basic arrangement of the sensor is illustrated with the analog voltage output, which needs to be A/D converted for next digital processing. Some sensors include the A/D converter on the chip.

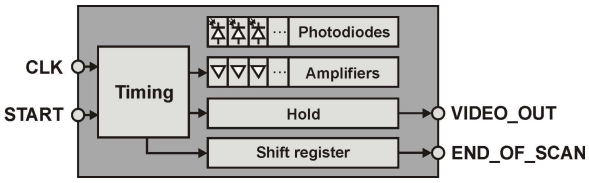


Figure 2. The structure and typical interface of line-scan CMOS sensor.

CMOS image sensor technology is an alternative to the more advanced image sensor technology based on the CCD (Charge-Coupled Device) principle. Although in broad terms is a function of CMOS and CCD technology similar, both technologies show in many respects rather different characteristics, which for certain types of applications one or the other technology favors.

CCD and CMOS technologies are so far regarded as largely complementary, it is estimated that this will be valid for a period of approximately one decade [11].

Approximate comparison of the CCD and CMOS sensors properties is shown in Table. 1 (available in [6, 9]).

TABLE I. COMPARISON OF CHARACTERISTICS OF CCD AND CMOS SENSORS

Feature	CCD Sensor	CMOS Sensor
Reading	One amplifier per a chip	One amplifier per a column
Noise	low	high
Reading speed	low	high
Dynamic range	high	low
Special features	possibility of precise control and a special exposure control (shuttering, analog convolution	windowing, access to each pixel, antiblooming
Supporting properties	off the chip	on chip
Control	complex (multi-phase clock signal, etc.)	simple
Price	high	low

IV. DESIGN OF CMOS LINE SENSOR CONTROLLER

The aim was to design the control system of line-scan sensor to obtain image data of the scene. In the previous chapter, there was shown a typical structure and control of line-scan sensors, it implies that the primary functions of the scanner include:

• generating the clock signal

• reading and converting the image information to digital data

• exposure time control by generating proper timing control signals

• generate control signals for configuring the sensor regarding to the requirements of a host system or operating conditions.

From the perspective of higher functionality of the system must also provide:

• the possibility of storing data captured during the off-line mode

• the possibility of communication with the host system for data transmission and system control.

Other features of the system may include a variety of application-dependent functions such as filtering the image data, scene change detection, calculation the centre of gravity of an image, edge detection and segmentation, histogram calculation and others. Implementation of these features, however, is beyond the scope of this work.

Hardware implementation of the proposed system was chosen with respect to the anticipated complexity of the system and as a suitable type was selected a Spartan-3 family FPGA. Availability and cost of

development boards on the market led to the Digilent's Spartan-3E development board Nexys2[12].

V. TIMING AND SYNCHRONIZATION

Image sensors are typically clocked by single clock signal, from which are derived next functions like the reading of image data, exposure control, and others. The operation of the sensor is then synchronous, which is well compatible with the control system design in FPGA, where function of the device is synchronized by one or more clock signals and where a synchronous design is usually used. With the anticipated connection or an asynchronous communication interface controls, however, raises the issue of timing signal to the synchronous input that can not be changed at arbitrary times, is to be secured properly. The system is therefore designed a method of synchronization signals at the interface of the synchronous and asynchronous system.

The developed system is expected to connect to a host system that will process the image data from the sensor. This provides some flexibility in deciding which of the presumed functions of the system will be implemented in the control system with FPGA and which will be implemented in software on the host system. This applies in particular to the type of superstructure functions like processing and image analysis, pattern recognition and data archiving. In the other hand, the functions such as waveform generation and timing control signals, control the exposure time, storing scanned data in memory or basic image preprocessing is necessary to be implemented in the control system.

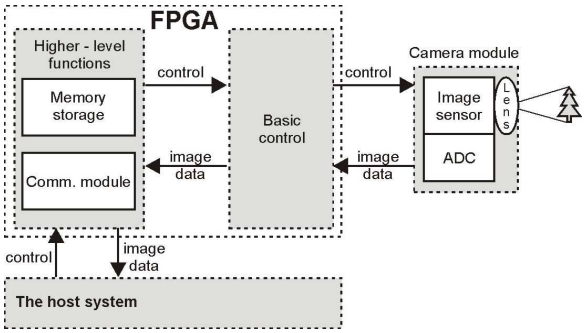


Figure 3. Block diagram of the designed system.

As mentioned in the introduction, a project partly builds on earlier experiments with a RPLIS2048 line-scan sensor and implemented construction of a prototype camera module with a series ARM7 microcontroller.

The reasons that led to this type of sensor in that camera prototype module are as follows:

- high number of pixels (2048),
- sufficiently high image reading frequency (2,5MHz),
- sufficient sensitivity (32 μV/electron),
- Dimensional compatibility with 1/3" flat sensors and the usability of standard lenses manufactured,
- availability of production samples,
- low price.

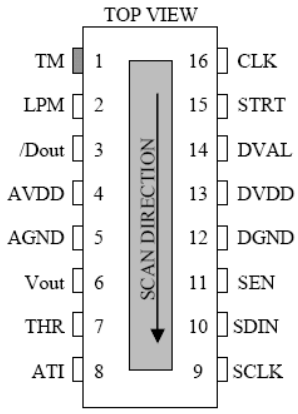


Figure 4. RPLIS2048 pinout [10].

VI. EXPERIMENTAL TESTS

During the tests was verified the functionality of the various settings available through the control registers and the functionality of reading image data.

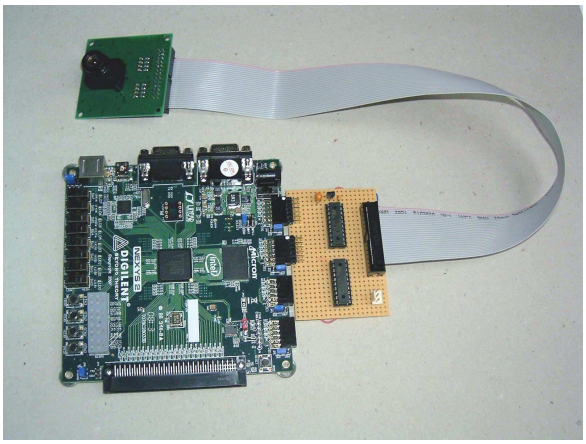


Figure 5. Camera module attached to the board Nexys2

Testing options were somewhat limited by the time of testing was not yet available software that would allow generation of appropriate sequences of commands and read the contents of system registers while maintaining temporal relationships.

Figure 6. Test pattern 1.

To test the function of camera module and sensor control system a test pattern was used, as shown in Figure 6. The test pattern was made of white, gray and black paper with a matte surface and it can therefore be regarded as a diffuse reflective. Figure 7 shows the voltage outputs of the sensor while scanning a test pattern. The system and data reading during tests were controlled by a TransPort utility, a part of the Digilent's Adept software [11].

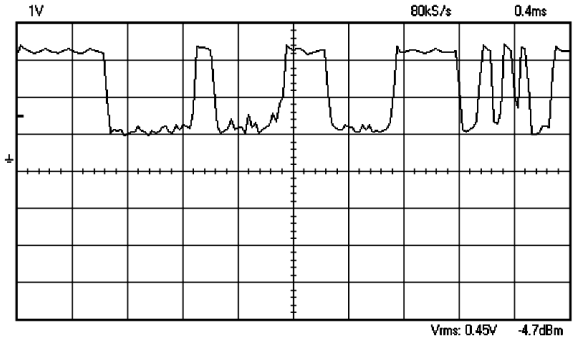


Figure 7. The voltage VOUT output of the sensor (test pattern 1).

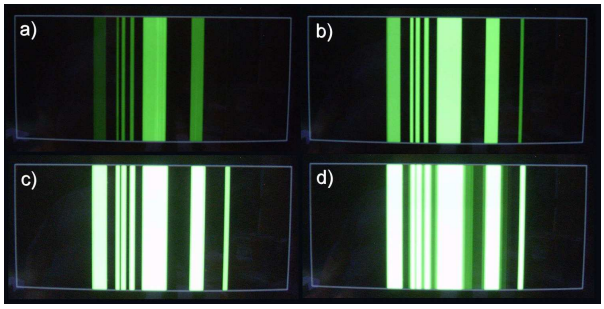


Figure 8. Scanning of test pattern in manual exposure mode. Settings: CamCtrl=0x80, SenCFG=0x00. a) Exp = 0x04 b) Exp = 0x08 c) Exp = 0x10 d) Exp = 0x20

VII. Conclusion

The hardware design consists of a camera module with a line-scan sensor with a resolution of 2048 pixels, the control part of the system was designed using VHDL and implemented in Xilinx Spartan-3E FPGA. For the prototype implementation was used Digilent's Nexys2 development board. The system is equipped with a USB interface for control and transmission of image data toward the host system.

The structure of the control system was designed to be easily customizable by adding new components. It is possible by use of VHDL, which can be used to create reusable components. Then, in future can be easily added components for filtering and correction, segmentation, motion detection and application-specific components for image processing. Structure of control registers is also designed to allow easy integration of optional components.

Partial functionality of the system was successfully tested. For the tests was used a universal TransPort utility by Digilent company, which allows easy access to data and system control registers through the USB interface.

There was successfully verified the basic functionality of the system, i.e. the general image acquisition, the possibility of acquiring a single image, change the bit depth of the image and save image data in memory for later retrieval. The automatic exposure based on keeping the average luminance of the frame works well for scenes of average brightness, too dark or too bright scenes can be captured with different exposure to the optimum in terms of image quality. Routine use of automatic exposure will require further refinement of the algorithm used.

Acknowledgment

The work and the contribution were supported by the project: Ministry of Education of the Czech Republic under Project 1M0567 “Centre of applied cybernetics”, student grant agency SV 4501141 “Biomedical engineering systems VII” and TACR TA01010632 “SCADA system for control and measurement of process in real time”. Also supported by project MSM6198910027 Consuming Computer Simulation and Optimization.

References

[1] Stankus, M., Penhaker, M., Kijonka, J., Grygarek, P. *Design and Application of Mobile Embedded Systems for Home Care Applications* In Proceedings of 2010 Second International Conference on Computer Engineering and Applications. ICCEA 2010, 19. – 21. March 2010, Bali Island, Indonesia, Volume 1NJ. IEEE Conference Publishing Services, 2010 p. 412-416. ISBN 978-0-7695-3982-9, DOI: 10.1109/ICCEA.2010.86

[2] Krejcar, O. *Large Multimedia Artifacts Prebuffering in Mobile Information Systems as Location Context Awareness*. In 4th International Symposium on Wireless Pervasive Computing, ISWPC 2009, February 11-13, 2009, Melbourne, Australia. pp 216-220. ISBN 978-1-4244-4299-7, DOI 10.1109/ISWPC.2009.4800591

[3] Labza, Z., Penhaker, M., Augustynek, M., Korpas, D. *Verification of Set Up Dual-Chamber Pacemaker Electrical Parameters.* In 2010 The 2nd International Conference on Telecom Technology and Applications, ICTTA 2010. March 19-21, 2010, Bali Island, Indonesia, Volume 2NJ. IEEE Conference Publishing Services 2010, p. 168–172. ISBN 978-0-7695-3982-9., DOI: 10.1109/ICCEA.2010.187

[4] Huang, F., Wei, S.K., et al. *Rotating Line Cameras: Epipolar Geometry and Spatial Sampling* [online]. University of Minnesota, March 2006 [cit. 2009-04-05]. Available from: <http://www.ima.umn.edu/preprints/mar2006/2105.pdf>.

[5] *Panoscan Panoramic Camera for High Speed Digital Capture* [online]. [cit. 2009-03-10]. Available from:<http://www.panoscan.com>.

[6] Donati, S. *Electro-Optical Instrumentation: Sensing and Measuring with Lasers*. Prentice Hall, 2004. ISBN 0-13-061610-9.

[7] Ohta, J. *Smart CMOS Image Sensors and Applications*. Boca Raton: CRC Press, 2007. ISBN 0-8493-3681-3.

[8] *Basics of Spectral Measurement* [online]. JETI Technische Instrumente GmbH, May 2005 [cit. 2008-04-05]. Available from<http://www.jeti.com/down/download.php?path=basics&file=basics.pdf>.

[9] Litwiller, D. *CCD vs. CMOS: Facts and Fiction* [online]. Dalsa, 2001 [cit. 2009-03-20]. Available from: <http://www.dalsa.com/public/corp/Photonics_Spectra_CCDvsCMOS_Litwiller.pdf>.

[10] *RPLIS-2048* [online].Datasheet.Panavision Imaging, 2006 [cit.2008-09-20]. Available from: <http://www.panavisionimaging.com/PDF/PDS0012RevM.pdf>.

[11] *Adept 1.10* [online].Software.Digilent Inc. [cit. 2008-12-10]. Available from <http://www.digilentinc.com/Data/Products/ADEPT/DASV1-10-0.msi>.

[12] *Nexys2 reference manual* [online].Datasheet.Digilent Inc., 2007 [cit.2008-10-11]. Available from <http://www.digilentinc.com/Data/Products/NEXYS2/Nexys2rm.pdf>

Simulation Tool for Playing Gaming Machines

Bohdan Borowik
Electrical & Automation Department
University of Bielsko-Biala,
Bielsko-Biala, Poland
bo@borowik.info

Miroslav Voznak
Dept. of Telecommunications
VSB – Technical University of Ostrava
Czech Republic
miroslav.voznak@vsb.cz

Abstract: **The paper presents the system designed and tested by the author for supervision and control of the game machine operations and for graphical presenting the course of the game. The hardware device, connected to the START signal line allows for initiation of 8192 playings (draws) and also for reading every time the total-in and total-out values from the PLAY and WIN meters. Acquired data is then processed with the software created by the author and results are presented graphically.**

Keywords-Komponent*: *playing gaming machines, microcntroller, PIC

I. INTRODUCTION

The paper presents the system designed and tested by the author for supervision and control of the game machine operations and for graphically presenting the course of the game.

The hardware device, connected to the START signal line allows for initiation of 8192 playings (draws) and also for reading every time the total-in and total-out values from the PLAY and WIN meters.

II. PRINCIPLE OF SYSTEM OPERATION

At starting time user chooses the bet from the set of available values. Then system sets the frequency of draws, the biger the value of bet, the lower draws frequency (bigger time interval between draws), according to table 1.

Table 1

Bet [credits]	time interval between draws [s]
5, 10, 20	1.8
50	2.2
100	3
200	5
500	7
1000	13

Thus the electro-mechanical meter for credited values (total-in) is able to increment its value with the bet.
Time interval between pressing and releasing START switch (duty cycle) has to be lower, than 0.8 s, otherwise the game enters auto-play mode.

For each drawing, the system preserves two values: drawing number and value of winning for this drawing. Those two values are shown on display as well as are sent through RS232 to terminal on PC computer for further processing.
When testing is over (max 8192 drawings) its results are presented graphically on PC.

Switches for stopping and resuming operation in any moment are also provided. Investigations were conducted with the popular gaming machine Eurogame Novomatic. The games installed were: Burning Hot, Ultra Hot, Am Poker, Burning Target, and Sizling Hot. Fig. 1 shows block diagram of the system.

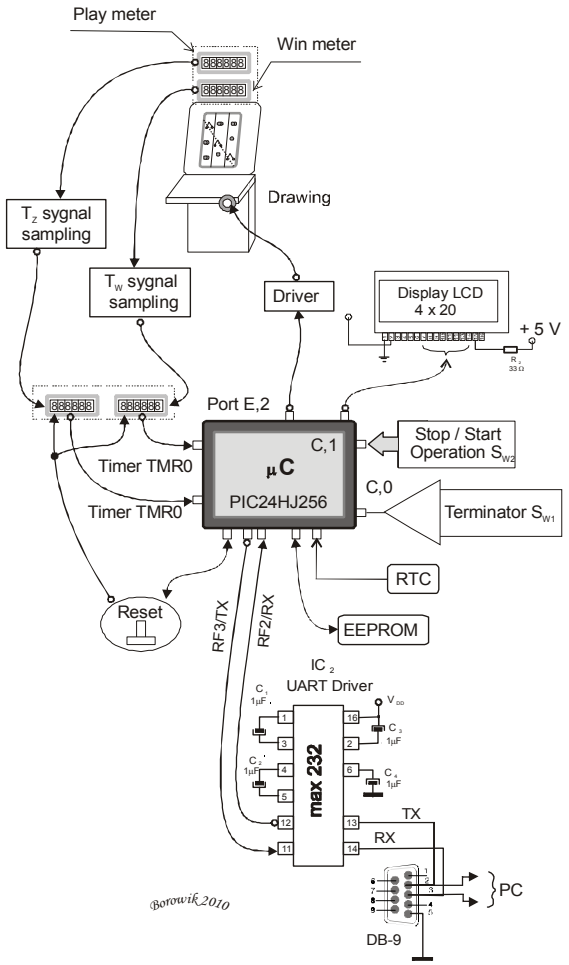


Figure. 1. Block diagram of the game simulation system working with the Eurogame Novomatic machine

Repetitiveness in popular games is of the order of 10 000. We designed making great number of drawing (<8192), so that results are representative for the game.
We do not take into account bonus play., f.ex. when playing Burning Target. When the bonus option appears, the game operation simply omits this option. Thus we obtain data comparable among different games. Option "play within the play" is not possible.

Such nested structure, like binary tree would have been trap making final decision – undeterminable. The system should be transparent therefore options of bonus play and "play within the play" are not taken into account.

III. SYSTEM DESCRIPTION

The heart of the presented system is the microcontroler PIC24HJ256. The system enter 2 input signals from the gaming machine: signal Tz from the PLAY meter and signal Tw from WIN meter. The first of them inputs to TMR2 module of the PIC microcontroler (pin 6). The second:
Signal Tw enters TMR1 (pin 74) of the microcontroler.

The system has 2 switches:
Sw1 -Start switch connected with C0.
Sw2 (toggle) connected to C1. After pressing it, we can stop/resume the simulation at any moment.

After pressing the switch SW1 the game simulation begins. In defined time intervals other signals (Sw) are released that simulate pressing PLAY button. Time intervals depend on the bet value according to table 1.
Waveform of the signal is shown in fig.2

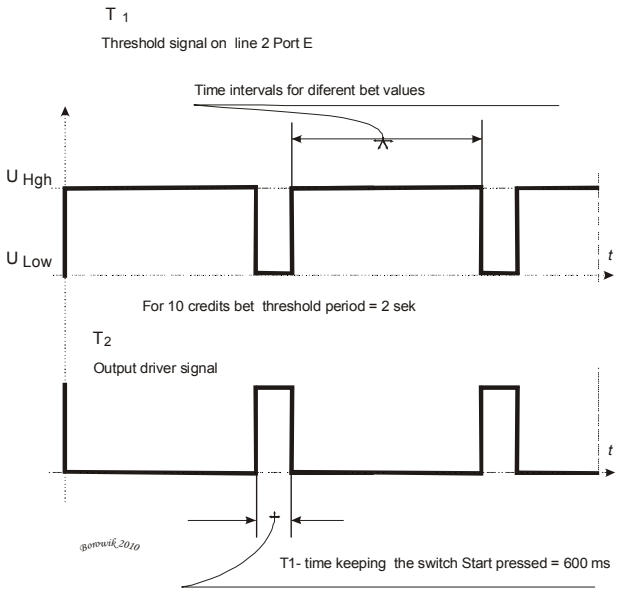


Figure. 2. Waveform of the signal that simulate game playing.

When winning is received, then Sw signal becomes low and stops further drawing. Simultaneously signal Tw from the WIN meter enters to TMR1 timer. It continues until total

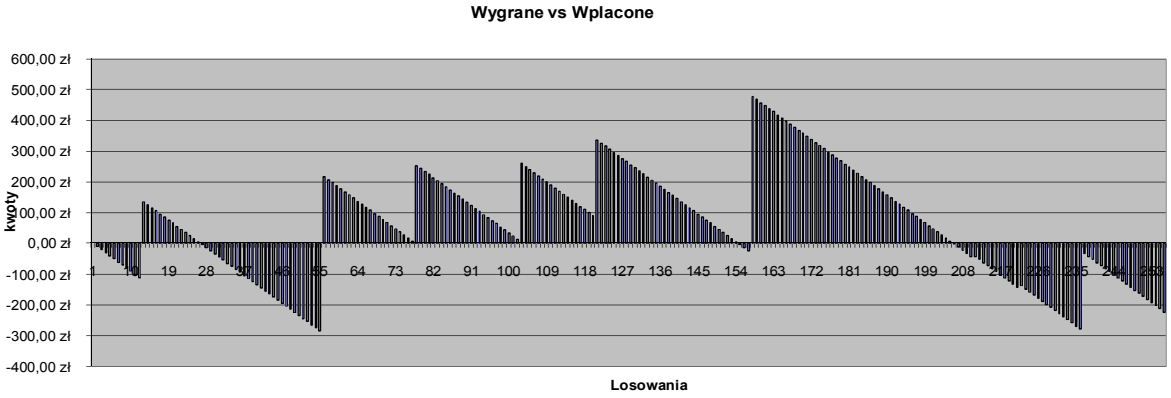


Fig. 5 Winnings, credited values and cumulated amounts against ordinal number of drawing

I. REFERENCES

[1] B. Borowik ,Interfacing PIC Microcontrollers to Peripherial Devices" Springer 2011

[2] B. Borowiki, I. Kurytnik, „Mikrokontroler PIC w zastosowaniach," PAK, Warszawa, 2009 pp 68-145

[3] C. Hudlestone, „Inteligent Sensors Design using the Microchip dsPIC," Elsevier, New York, 2007 pp130-160

Recent Advances in Networks and Communications

chairman:

Igor Piotr KURYTNIK

Method of attacks visualization in wireless sensor networks

Bohdan Borowik, Igor Piotr Kurytnik
Department of Electrical Engineering and Automation
University of Bielsko-Biala
Bielsko-Biala, Poland
bborowik@ath.bielsko.pl, ikurytnik@ath.bielsko.pl

Volodymyr Karpinskyi
Department of Electrical Engineering Power Consumption Systems
Ternopil Ivan Pul'uj National Technical University
Ternopil, Ukraine
vkarpinskyi@gmail.com

Abstract—**The paper proposes the method of attacks visualization in wireless sensor networks that reduces distance measuring error between sensors by mixed noise described by Bessel functions of imaginary zero and higher order argument. The method also increases the accuracy of the reconstructed smoothed topological sensor network surface and extends the functionality of providing the analysis of visualized shape of a reconstructed smoothed topological sensor network surface.**

Keywords-visualization; WSN (Wireless Sensor Network); attack; security

I. WAYS OF ANALYSIS OF MODELING A VISUALIZATION OF WORMHOLE ATTACKS DETECTION IN A WIRELESS SENSOR NETWORK

Considering ever more increasing cases of attacks realization on Wireless Sensor Networks (WSNs), including menacing the Wormhole attack (taking into account various fields, such as IT, industrial and specialized network systems, power and pipeline maintenance, ecology, agriculture, supermarkets and other strategic and financial powerful institutions), there is a big need for the protection from such attacks. One of the approaches is using of a simulation of a visualization of a Wormhole attack detection.

A well-known visualization method of the environment in the data center, using a WSN, consists of the performance of the following procedures [1]:

- connection establishing with network sensors of ecological monitoring deployed in a specific one of many points of object monitoring area.
- periodical values collection of measured parameters from each environmental network sensor of ecology monitoring made by the WSN,
- calculation by the extrapolation method the environmental variables at object's intermediate points based on the collected values of the environment measured parameters in specific sections,
- visualization of the environment state on the object based on measured parameters values and calculated values of ecological values.

The disadvantage of this method is the limited security opportunities of performed attacks on wireless sensor networks, caused by the fact that this method does not allow to visualize detection of the wormhole attack in a wireless sensory network.

Another way of modeling WSN topology design to protect networks from attacks involves the followig evaluations [2]:

- it's security degree against threats of that network,
- parameters of the network topology design depending on the security degree,

and these reduce to:

- network design depending on the function at least one design parameter and/or at least one threat parameter
- assessment of the model, changing at least one design parameter to determine the effect on a one threat parameter
- selection of one designed feature based on the assessment by which it was synthesized a reasonably low value of at least one threat parameter.

The disadvantage of this method is that it covers only a preliminary distribution of keys among the WSN nodes, the assessment of common key distribution is between a pair of nodes, a mutual verification of the shared key with a node pair establishes the degree of authenticity and additionally it provides nodes distribution in accordance with design parameters that only partially protect network and in a result is inadequate to detect wormhole attacks. In addition, the method applies only to the design stage of the WSN, thus has the limited application, especially in case of network exploitation and wormhole detection during its exploitation.

A part of publications is devoted to resolve the problem of remote pipelines (water, sewer, etc.) monitoring and power lines of utilities monitoring using sensor networks and a simulation of an operation and construction modes of a WSN for these tasks [3-7]. Specifically, mentioned there way of monitoring, is to use the WSN that contains [3-5]:

- sensors deployed for monitoring pipelines and power lines,
- sensors for receiving data, related to pipelines and power lines, and for communicating with sensor data,
- remote sensor interface, which contains the communication device that receives data transmitted by sensors and further transmits it,
- communication devices that accept data transmitted by remote sensor interface and transmit data directly or indirectly outside of the WSN.

These decisions are characterized by the lowest level of a WSN protection against attacks caused by the lack of security and the inability to visualize worm attacks.

According to the approach proposed in [8], this model provides multiple ways of wormholes, because it creates an interconnect structure, which effectively uses the presence of a silicon and a wormhole flow of messages operations and the temporary blocking of wormhole message is carried through a signal control.

However, the disadvantage of this model is that despite the temporary blocking message of wormhole through a control signal, the message has always a free path available for the continued progression of the wormhole via the bus through the traffic down in another column from the input port to the output port.

Moreover, the method of its realization has a limited usage - only in hardware implementations of the interconnection family structures for scalability, small delay and chip performance, which supports a wide range of applications, including supercomputer interconnections, LAN, IP and ATM switches, central telephone switching station, video servers on demand, routes to central database servers, high-speed line work stations, and many others, with no way to detect wormhole attack in wireless sensor networks through its visualization.

Recently there has been posted a publication in which the use of a visualization modeling was proposed to detect attacks on WSNs. In [9] the method and the visualization modeling of wormhole attacks on WSN, consist of the following steps: measuring the distance between the sensors based on the power level of received signal provided by the error modeling of distance measurement simplified by a uniform noise; Dijkstra's algorithm for calculating the shortest distance between each pair of sensor and building a matrix of distances in the network; WSN topological surface reconstruction by multidimensional scaling and calculating the virtual position of each sensor; smoothing the reconstructed WSN topological surface by approximation of plane for each sensor based on its coordinates and neighbors coordinates with further determining of the new sensor position using the previous coordinates and its projection on the approximated plane; and analysis of smoothed shape analysis of reconstructed WSN topological surface and identifying fake sensor neighbor connections caused by wormhole attack.

To the proposed method and model disadvantages we can include the usage of the simplified modeling error of distance measuring between sensors by a homogeneous noise, that has a partial limited nature and does not allow to take into account the real noises in WSN, due to a significant distance error measurement between sensors, insufficient precision of the WSN topology reconstructed surface smoothing and limited functionality analysis of visualized shape of the reconstructed smoothed surface of sensor network.

Therefore there is a task problem of developing the modeling visualization to detect wormhole attacks on WSN with reduced error of distance measurement between sensors, higher precision smoothing of the reconstructed WSN topological surface and enhanced functionality analysis of visualized shape of smoothed reconstructed WSN topological surface.

II. MODELING METOD OF THE VISUALIZATION OF THE WORMHOLE ATTACK DETECTION ON A WIRELESS SENSOR NETWORK

The proposed method applies to a WSN, includes its security from a wormhole attack by a visualization and is designed for a usage in various industries, such as IT, industrial and specialized network systems, power industry, pipeline maintenance, ecology, agriculture, supermarkets, institutions and establishments of strategic financial and important purpose, etc. The goal of developing a visualization modeling method of a wormhole attacks detection on a WSN lies in:

• reducing a distance error measurement between sensors is done by providing its modeling of mixed noise, that described by Bessel functions of imaginary zero and higher order argument;

• improving accuracy smoothing of reconstructed WSN topological surface,

• expanding functionality of visualized shape of smoothed reconstructed WSN topological surface,

Solving these tasks is achieved through consistent implementation of the following procedures:

• distance measuring between sensors based on the power level of received signal,

• reconstruction of the WSN topology surface by multidimensional scaling and calculating the virtual position of each sensor,

• smoothing of reconstructed WSN topological surface,

• analysis of visualized shape of smoothed reconstructed WSN topological surface and detection of fake sensor neighbor connections caused by wormhole attack.

Thus for:

• distance measuring between sensors is used distance measurement error modeling between sensors by mixed noise, described by Bessel functions of imaginary zero and higher order argument (Fig. 1),

• smoothing the reconstructed WSN topological surface using the Delaunay Triangulation and kriging interpolation (Fig. 2),

• analysis of visualized shape of smoothed reconstructed WSN topological surface is applied the triangle mesh with sensors placed in their vertices with coordinates in Euclidean space (Fig. 3). Positions in Fig. 1-3 marked by 1, 2, 3, 4 - WSN sensors.

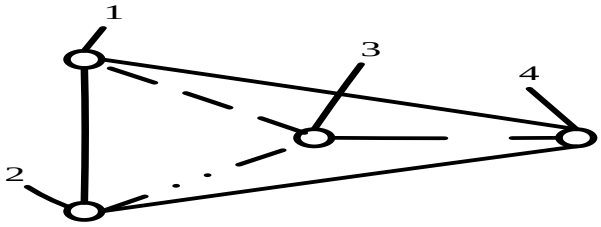


Figure 1. Scheme of distance measuring between sensors based on power level of received signal

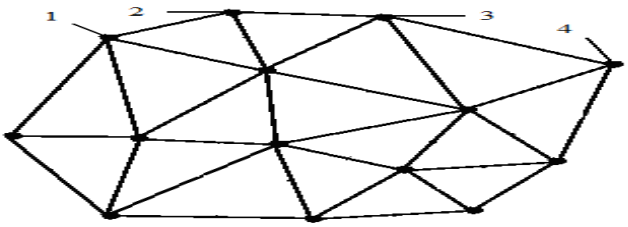


Figure 2. Delaunay Triangulation

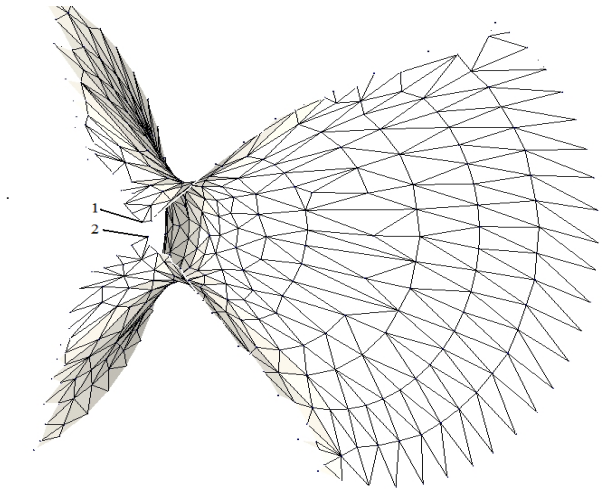


Figure 3. The modeling vizualized wormhole detection in WSN based on kriging interpolation of the reconstructed topological network surface and triangle mesh with sensors placed in their vertices with coordinates in Euclidean space

The method reduces distance measurement between sensors based on the power level of the received signal is conducted (Fig. 1). Then the modeling error of distance measuring is carried between sensors by mixed noise, described by Bessel functions of imaginary zero and higher order argument:

The next step lies in rectification of the measured distance between the sensors based on power level of the received signal:

Further reconstruction of the WSN topological surface is carried by a multidimensional scaling and calculating of the virtual position of each sensor:

Then the reconstructed WSN topological surface is made more smooth.

First is conducted the Delaunay triangulation (Fig. 2). Then a kriging interpolation of the reconstructed topological surface is performed:

Then is conducted: a) analysis of the visualized shape of the smoothed reconstructed WSN topological surface triangle mesh with sensors placed in their vertices with coordinates in Euclidean space, b) the detection of fake connections between neighbor sensors caused by wormhole attack (Fig. 3). On the basis of the results of the analysis it can be stated that a wormhole attack bend of smoothed reconstructed WSN topological surface, attracting sensors to each other and creates fake connection at points 1 and 2 of sensors position (Fig. 3).

III. CONCLUSIONS

The proposed method can reduce the distance error measuring between sensors, improves the accuracy of the smoothed reconstructed WSN topological surface and extends the functionality analysis of the visualized shape of smoothed reconstructed WSN topological surface.

REFERENCES

[1] Apparatus and method for visualizing environmental conditions in a data center using wireless sensor networks : US Patent 702/188 : Int.Cl. G06F15/00 / Ramin Y., Pandey R. – US20100280796A1 ; Pub. Date. 04.11.10. – 19 p.

[2] Modeling a sensor network design to secure a network against attack : US Patent 380/278; 340/539.22; 370/338 : Int.Cl. H04L9/00 / Roy S. S. R., Mukhopadhyay D., Thejaswi PS C. – US007804962B2; Pub. Date. 28.09.10. – 22 p.

[3] Remote monitoring of pipelines using wireless sensor network : US Patent 73/49.1; 73/40.5 R; 73/86; 73/865.9; 702/113 : Int.Cl. G01M3/28; G01N17/00 / Sabata A., Brossia S. – US007526944B2; Pub. Date. 05.05.09. – 6 p.

[4] Sensor networks for pipeline monitoring : US Patent 340/870.07; 340/854.5; 73/40.5 R; 73/49.1 : Int.Cl. H04Q5/00 / Twitchell R.W. Jr. – US007830273B2; Pub. Date. 09.11.10. – 13 p.

[5] Sensor networks for monitoring pipelines and power lines : US Patent 340/870.01; 370/252; 73/49.1 : Int.Cl. G08C17/00 / Twitchell R.W. Jr. – US007705747B2; Pub. Date. 27.04.10. – 13 p.

[6] Sensor network system : US Patent 455/412.1; 455/456.1; 455/435.1 : Int.Cl. H04L12/58 / Kato H., Miyao T. – US007680486B2; Pub. Date. 16.03.10. – 24 p.

[7] System and program product for signal transmission between a sensor and a controller in a wireless sensor network : US Patent 340/539.1; 340/870.01; 343/757; 455/69 : Int.Cl. G08B1/08 / Coronel P. E., Furrer S.; Shott W. H. – US007782188B2; Pub. Date. 24.08.10. – 16 p.

[8] Multiple-path wormhole interconnect : US Patent 340/870.01; 370/252; 73/49.1 : Int.Cl. H04Q 11/00; H04L12/28; G06F15/173 / Hesse J. E. – US007382775B2; Pub. Date. 03.06.08. – 93 p.

[9] W. Wang, B. Bhargava "Visualization of Wormholes in Sensor Networks," Workshop on Wireless Security 2004 : Proceedings of ACM Workshop WiSE'04, October 1, 2004, Philadelphia, PA USA. – Pp. 826-838.

Research of realization a distributed attacks in computer network

Mikolaj Karpinski
Department of El. Eng. and Automation
University of Bielsko-Biala
Bielsko-Biala, Poland
mkarpinski@ath.bielsko.pl

Yaroslav Kinakh
Department of Computer Engineering
Ternopil National Economic University
Ternopil, Ukraine
kinakh@lycos.com

Uliana Yatsykovska
Dpt. of Computer Science
Ternopil Ivan Pul'uj
National Technical University
Ternopil, Ukraine
price@ukr.net

Abstract— **The paper proposes a mathematical model of communication of client server that includes the probability of compromised node and the number of all possible routes that can have an admission to access points, have done a comparative characteristics of attacks DoS / DDoS in computer network.**

Keywords- computer network, denial of service, mathematical model

I. INTRODUCTION

Because the principle of open networks and access to them are specific features of their structure and processes of operation, such as openness, protection [1], characterized by significant heterogeneity [2]. At present, special attention focuses on new areas of development and improvement of data networks. Among them should provide wireless (mobile) networks. Such networks provide the user with unique opportunities for fast access to remote network resources, including the global network Internet, limiting his mobility, not linking to the wired communication lines [3].

With the development and complication of the tools, techniques and processes of information processing increases dependence of modern society on the degree of security used his information technology [4].

Computer network providing every opportunity for exchanging data between the client and server, but now widely distributed attack denial of service clients, the determination of distributed attacks in the network is particularly acute. The most common types of such attacks are DoS / DDoS attacks, which deny certain users of computer network services.

With the constant development of computer networks and the increasing number of users grows and the number of new types of attacks to denial of service. DoS / DDoS attacks are characterized by a straightforward implementation complexity and resistance, which poses new problems of researchers, who are still not yet resolved. Analysis of recent publications shows that exercise is accompanied by attacks: interception of confidential information to unauthorized use of network bandwidth and computational resources, the spread of false information, violation of network administration.

II. THE MAIN PART

To solve the task should use the classification of information threats, DoS / DDoS attacks and formalized models [1] measure the impact on job performance computer network.

This will effectively solve the problem of detecting attacks on computer network access point. Construct formal mathematical models of probability of information threats, DoS / DDoS attacks based on the linear form of the method of weighting coefficients.

$$P\_^2T(P) = \alpha_1 P_{\text{Confidenti ality}} + \alpha_2 P_{\text{Integrity}} + \alpha_3 P_{\text{Accessibility}} \quad (1)$$

$$P\_DoS(P) = \beta_1 P_{Smurf} + \beta_2 P_{Fraggle} + \beta_3 P_{SYNFlood} + \beta_1 P_{DNS} \quad (2)$$

$$P\_DDoS(P) = \delta_1 P_{Trinoo} + \delta_2 P_{TAN/TF2K} + \delta_3 P_{Stacheldraht} \quad (3)$$

where P_IT_(P) – probability of information threats;
P_DoS(P) – probability of DoS attacks;
P_DDoS(P) – probability of DDoS attacks;
α_i – weighting coefficients, where $\alpha_i \in [0;1]$;
$\beta_з$ – weighting coefficients, where $\beta_з \in [0;1]$;
$\delta_з$ – weighting coefficients, where $\delta_з \in [0;1]$.

The mathematical model defining the matrix network activity, according to which make conclusions about the realization of attack:

$$\alpha_{IT} = \begin{vmatrix} \alpha_1^a & \alpha_2^a & \alpha_3^a \\ \alpha_1^b & \alpha_2^b & \alpha_3^b \\ \alpha_1^c & \alpha_2^c & \alpha_3^c \\ \alpha_1^d & \alpha_2^d & \alpha_3^d \\ \alpha_1^e & \alpha_2^e & \alpha_3^e \\ \alpha_1^f & \alpha_2^f & \alpha_3^f \\ \alpha_1^g & \alpha_2^g & \alpha g \end{vmatrix}, \beta_{DoS} = \begin{vmatrix} \beta_1^a & \beta_2^a & \beta_3^a & \beta_4^a \\ \beta_1^b & \beta_2^b & \beta_3^b & \beta_4^b \\ \beta_1^c & \beta_2^c & \beta_3^c & \beta_4^c \\ \beta_1^d & \beta_2^d & \beta_3^d & \beta_4^d \\ \beta_1^e & \beta_2^e & \beta_3^e & \beta_4^e \\ \beta_1^f & \beta_2^f & \beta_3^f & \beta_4^f \\ \beta_1^g & \beta_2^g & \beta_3^g & \beta_4^g \end{vmatrix},$$

$$\delta_{DDoS} = \begin{vmatrix} \delta_1^a & \delta_2^a & \delta_3^a \\ \delta_1^b & \delta_2^b & \delta_3^b \\ \delta_1^c & \delta_2^c & \delta_3^c \\ \delta_1^d & \delta_2^d & \delta_3^d \\ \delta_1^e & \delta_2^e & \delta_3^e \\ \delta_1^f & \delta_2^f & \delta_3^f \\ \delta_1^g & \delta_2^g & \delta_3^g \end{vmatrix} \quad (4)$$

The research has shown that all types of attacks evenly affecting computer network. With increasing probability kinds of attacks the probability of information threats and DoS/DDoS attack increases directly proportional. The denial of service attack has the greatest impact on network performance. But to discern what kind of attack is practically implemented, these models do not allow.

To determine the types of attack that is implemented, form the mathematical model of communication and customer service, which includes the likelihood compromise node and the number of ways to whatever they access points.

$$\text{I} \quad \alpha_i^a = \frac{1}{k}\left[P_{AP}^1 + P_{AP}^2\right]$$

$$\text{II} \quad \alpha_i^b = \frac{1}{k}\left[2P_{AP}^1 + P_{AP}^2\right]$$

$$\text{III} \quad \alpha_i^c = \frac{1}{k}\left[P_{AP}^1 + 2P_{AP}^2\right]$$

$$\text{IV} \quad \alpha_i^d = \frac{1}{k}\left[2P_{AP}^1 + 2P_{AP}^2\right] \quad (5)$$

$$\text{V} \quad \alpha_i^e = \frac{1}{k}\left[2P_{AP}^1 + P_{AP}^2\right]$$

$$\text{VI} \quad \alpha_i^f = \frac{1}{k}\left[P_{AP}^1 + 2P_{AP}^2\right]$$

$$\text{VII} \quad \alpha_i^g = \frac{1}{k}\left[P_{AP}^1 + P_{AP}^2\right],$$

where $\alpha_i^{\grave{a},b,c,d,e,f,g}$ – weighting coefficient,

a, b, c, d, e, f, g – model of communication,

i – types of attacks,

k – number of possible paths from AP to T.

Here are the results of numerical experiment with the model (5) in graphic form (picture 1). Legend:

α – weighting coefficient,

n – number of nodes,

$\sum_{i=1}^{n} P_{AP}^i$ – the total number of probably compromised access points.

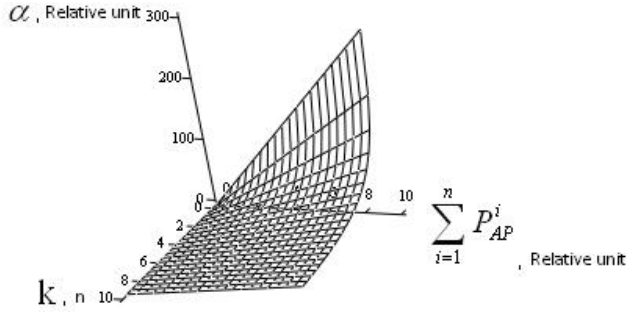


Figure 1. Dependence of probability weights compromised access points, and number of whatever routs

The research have shown that as the number of ways to whatever they can from client to server network activity is low, so the practical realization of attack is difficult to determine. For small values k, the active of network is growing rapidly, the attack is determined unambiguously. Level of the compromised nodes has a little impact on network activity in general, since these units do not determine the process routing.

To distinguish between that attacks was realized, we use Table I which analyzed the way to and through compromised node.

Table I. Comparative characteristics of attacks DoS/DDoS, computer network

Type of attack		**Route, k**			**Passing through the compromised node**
		min	***indefinite***	***determined***	
DoS	Smurf	-	+	-	-
	Fraggle	-	+	-	-
	SYN Flood	-	+	-	+
	DNS	-	-	+	-
DDoS	Trinoo	-	+	-	+
	TAN/ TF2K	-	-	+	+
	Stachel draht	-	+	-	+

It should be noted that the attacks and DNS TAN/TF2K implemented on a specific path, because in a computer network they are easy to detect by analyzing traffic. Traffic activity increases significantly in the implementation of such attacks. In other cases it is difficult to determine the type of threat.

III. CONCLUSIONS, RECOMMENDATIONS AND PERSPECTIVES FOR FUTURE DEVELOPMENT OF THIS AREA

Research have shown that the formal mathematical model of probability information of threats and DoS/DDoS attacks based on the linear form of the method of weighting coefficients do not allow to discern what kind of attack is practically implemented in a computer network, because with increasing probabilities of attack types increases directly proportional probability information of threats and attacks, DoS/DDoS.

Dependence of probability weights compromised access points and ways of whatever they have shown that for small values k active network is growing rapidly and clearly defined attack. When increasing the number of ways to whatever they can from the client to the server, practical realization of attack is difficult to determine because of the low activity of the network. Level nodes of compromise have a little impact on network activity in general, since these units do not determine the process routing.

REFERENCES

[1] Y. Mylokum, "Models and methods of providing services in secure computer networks", thesis dis. for the degree of candidate engineering science: spec. 05.13.05 Computer Systems & Components, Aviation University of Ukraine, Kyiv, 2009, 19 p.

[2] Chang Shu (China), "Adaptive method of traffic flow computer networks", thesis dis. for the degree of candidate engineering science: spec. 05.13.05 Computer Systems & Components , Aviation University of Ukraine, Kyiv, 2009, 20 p.

[3] A. B. Kolesnik, "Information technology and management tools for adaptive mobile wireless computer networks", thesis dis. for the degree of candidate engineering science: Spec. 05.13.06 "Information technology", Kharkov National University of Radio Electronics in Ukraine, Kharkov, 2008, 25 p.

[4] R.A. Hayrapetyan, "Methods for protecting software from unauthorized access and malicious programs", thesis dis. for the degree of candidate engineering science: Spec. 05.13.21 "System of information security", Odessa National Polytechnic University of Ukraine, Odessa, 2009, 18 p.

Regression mathematical model of subscriber extension line

Iva Petrikova
Department of Telecommunications
VSB–Technical University of Ostrava, Faculty of Electrical Engineering and Computer Science,
17. listopadu 15, 708 33 Ostrava – Poruba, Czech Republic
Iva.petrikova@vsb.cz

***Abstract* — The paper focuses on the definition of a regression mathematical model for metallic subscriber extension lines implemented using a symmetric pair cable. The regression model is compared with an analytical model based on a theoretical description of transfer parameters for this type of line. The output of the paper should demonstrate the impact of electromagnetic interference on the symmetric pair. The paper also describes the method to identify the interference sources and type using the regression mathematical model.**

Keywords – subscriber extension line, regression mathematical model of wiring, metallic wiring, attenuation, line input impedance.

I. INTRODUCTION

One of the issues currently being addressed in the area of telecommunications technology is how to enhance the quality of the services provided and/or how to maintain their quality. As for transport network, the solutions in this area are already rather sophisticated, yet access networks and subscriber lines are their weak spots.

As these networks are used to provide multimedia services and data transfers and the signal transferred is in a digital format, the transfer quality needs to be higher than that of analogue transfers. The services mentioned above require higher transfer speed and subsequently also a higher frequency band [5]. At present, metallic subscriber lines are the most commonly used; and metallic wiring is still being used to connect new subscribers. This is essential for the quality of the whole transfer chain, though in most cases it is only the last section, i.e. the line from the switch board that is metallic. Symmetrical pair cables ensure the connection. They are commonly used in telephone networks for analogue signal transfers. With this type of wiring, the biggest factors influencing the transfer quality are electromagnetic interferences caused by various external sources and the topology of the extension line itself. The complexity of the issue is due to the randomness of interference, different and frequently unknown route of a particular line, various detours, etc. All these factors have an adverse impact on the quality of the transferred signal, both analogue and digital.

Network operators such as Telefónica O2 admit that interference sources are searched for and identified randomly and intuitively.

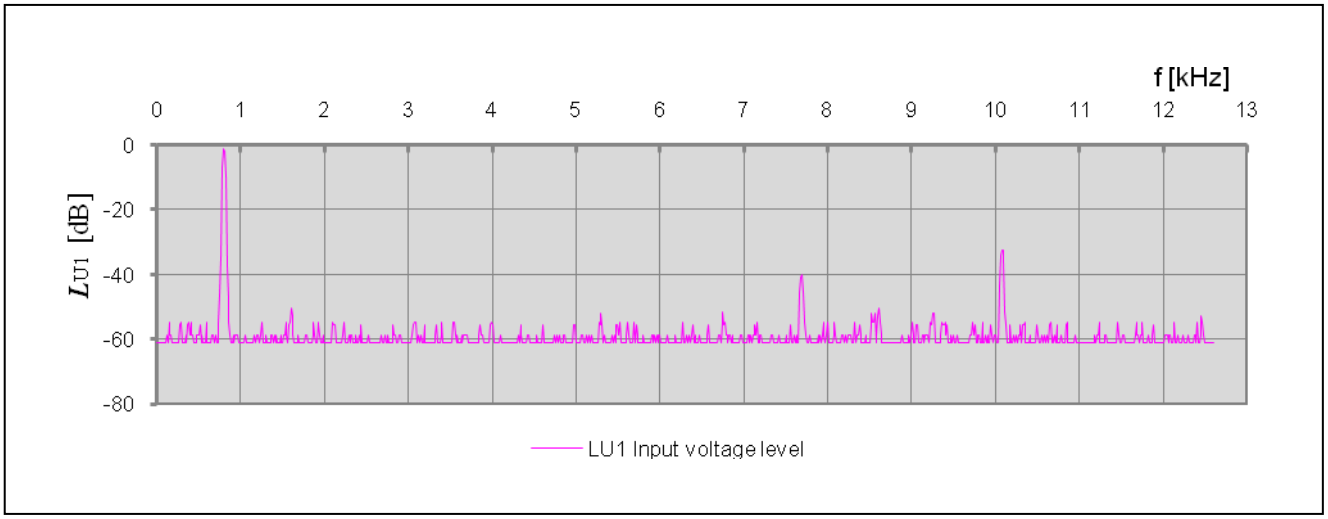


Figure 1. Input signal range

In many instances, they are not eliminated even by laying down a new extension line. One possible solution to mitigate this issue is based on the analysis of measured signal progress in a line subject to interference. The quality of the measurements performed is the key factor determining the

interpretation of progress measured and the correct interpretation of outputs.

II. RESULTS OF PRELIMINARY MEASUREMENTS

The results of the preliminary laboratory measurements, described in detail in [2], support the new approach and its usefulness for practical life. The measurements were performed in a spectral range, and as they were inhibited by certain systemic errors, they should only be regarded as preliminary measurements.

Figure 1 shows the development of voltage measured at the beginning of the line without any interfering element. The graph clearly indicates the basic frequency of 800 Hz. Yet two other significant frequencies appear, of approximately 7.7 kHz and 10 kHz. These two are not harmonious frequencies. Most probably, these are interfering signals from nearby appliances.

Two additional measurements were performed using electromagnetic relay as the interference sources, with their wiring being periodically broken by a circuit-opening contact. In the first case, the interference source (12 V accumulator) emitted direct-current supply voltage – tool 1. The values of voltage are displayed on Figure 2. In the second case, the interference was alternating current voltage of 230 V/50 Hz – tool 2. The values of voltage for this case are displayed on Figure 3.

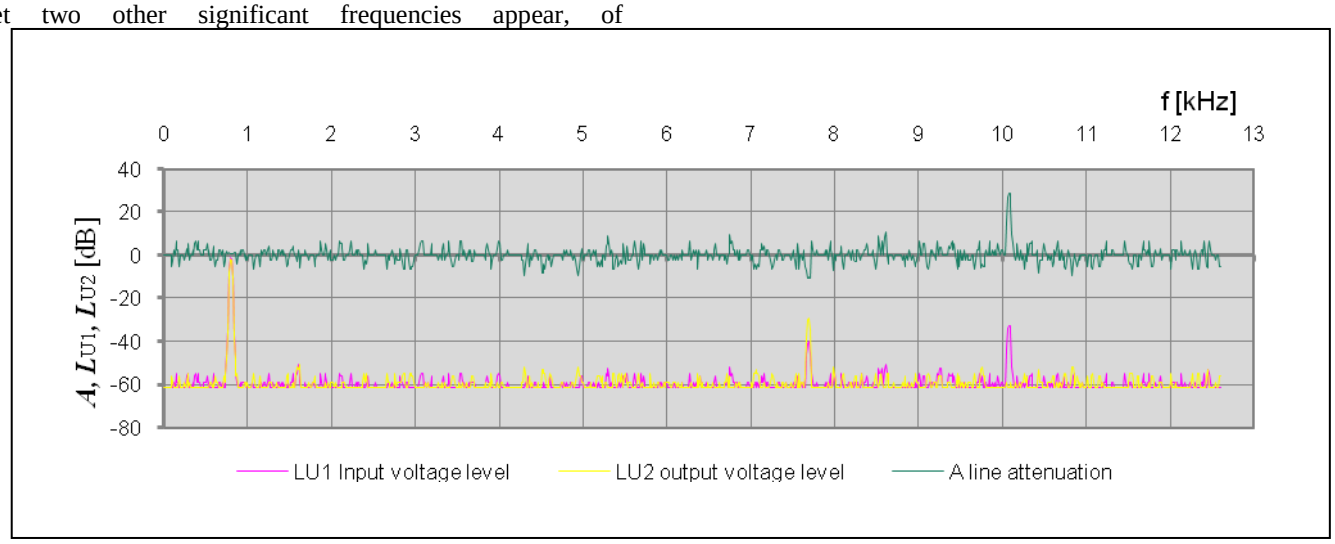


Figure 2. Voltage values when interfering with tool 1

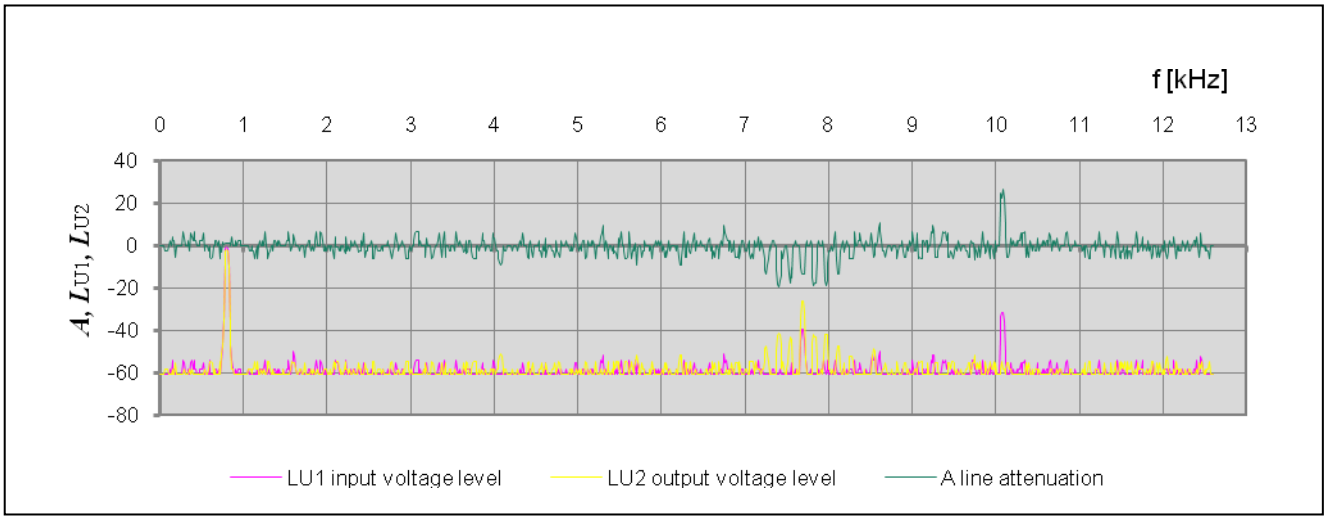


Figure 3. Voltage values when interfering with tool 2

Although the measurements are only preliminary, the results clearly show that the levels around frequency of 7 000 Hz increased in response to the interference. Figure 1 also shows the impact of the failure to adjust and unknown interference in the laboratory. These illustrative figures also fail to take into account the frequency relations of secondary line parameters.

The analysis performed indicates that the most suitable method to identify the source of the interference is that based on a mathematical model. It simulates the behaviour of a subscriber line, and its input variables are the standard measurable parameters characteristic for subscriber lines [3], [6]. Given the variety of conductors used and many possible forkings of subscriber extension lines, putting together a parametrical regression mathematical model seems to be the

optimum solution. Figure 4 below shows a schematic diagram of the model.

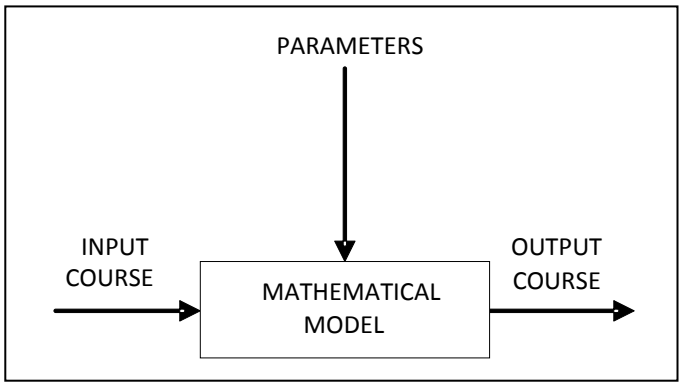


Figure 4. Schematic diagram of the mathematical model

III. REGRESSION MATHEMATICAL MODEL

The first draft of the output function of the mathematical model is as follows:

$$A = K \cdot \varphi_1(Z_2) \cdot \varphi_2(f), \tag{1}$$

where A stands for line attenuation, $\varphi_1(Z_2)$ function describes the impact of the load impedance of the Z_2 line, $\varphi_2(f)$ function describes the impact of frequency f and K is a constant [1].

The equation does not explicitly take into account the impact of the length of an extension line, yet its impact is included when calculated the line input impedance. Accordingly, further measurements were performed for the currently used SYKFY 2x2x0.5 cable of 250 m as described in Figure 5.

All measurements were performed using the AGILENT 33210A generator and the TDS 1002B oscilloscope. As the generator used has a non-symmetric output with only the real-part impedance of 50 Ω, it had to be connected through an adaptor. The same applies to the oscilloscope with a non-symmetric input and impedance of 1MΩ/20pF.

Nowadays, two transfer methods are being used at subscriber lines. The first being the analogue transfer, which is the most common one, and the second being the digital transfer. As for the digital transfers, transfer speeds of 64 kbit.s^{-1} and 2 048 kbit.s^{-1} are the most frequently used. A basic ISDN plug with transfer speed of 128 kbit.s^{-1} or 144 kbit.s^{-1} is also important. From the point of view of modelling, the crucial factor is the length of the impulse for the above stated transfer speeds. These are reflected into the impulse signal spectrum. To be able to put together a regression mathematical model and to analyse the measured values, it is sufficient to generate the signals periodically.

The complexity of the issue and the need to address it in the spectrum area makes the task extremely difficult, particularly in the experiment part.

Figures 6 and 7 display results for sinus type of the input signal with the frequency ranging from 300 to 3 400 Hz.

Figure 6 shows attenuation values obtained in measurements in relation to load impedance, with a constant frequency. Figure 7 shows attenuation values obtained in measurements in relation to load frequency with a constant load.

Based on these measurements, a regression mathematical model was developed. It determines attenuation as a function of the load impedance on the Z_2 line and as a function of frequency f.

The $\varphi_1(Z_2)$ function describing the impact of the load impedance on the Z_2 line is determined by a quadratic regression of values measured (see Figure 6) and can be determined as follows:

$$\varphi_1(Z_2) = 7 \cdot 10^{-7} Z_2^2 - 0{,}0016 Z_2 + 1{,}448 . \tag{2}$$

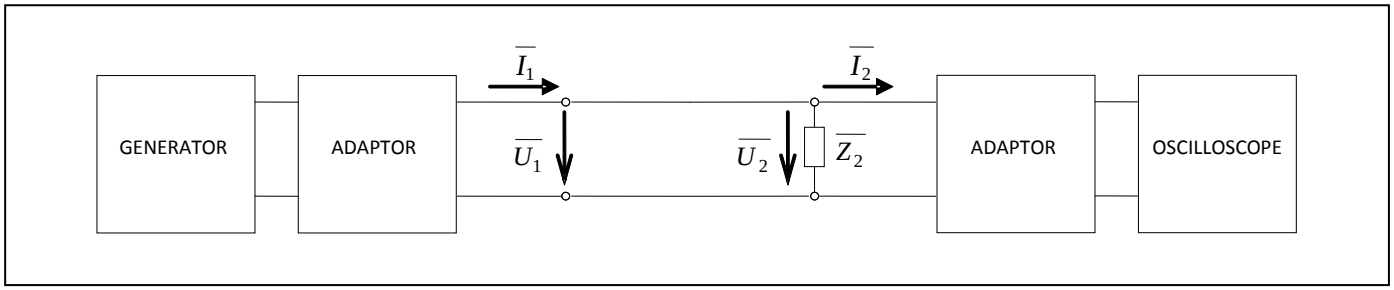


Figure 5. Schematic diagram of the experiment

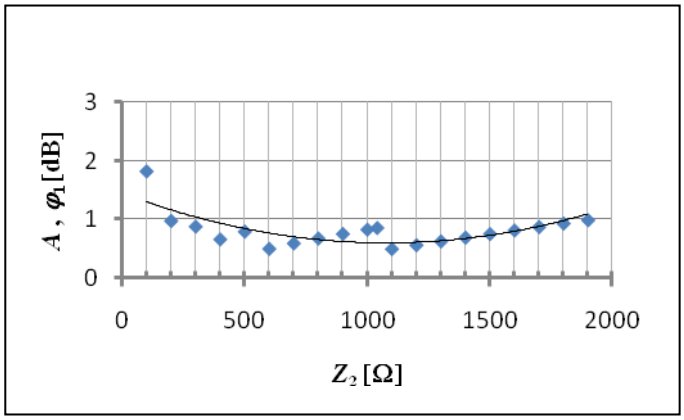


Figure 6. Relation between attenuation and load impedance for f = 800 Hz

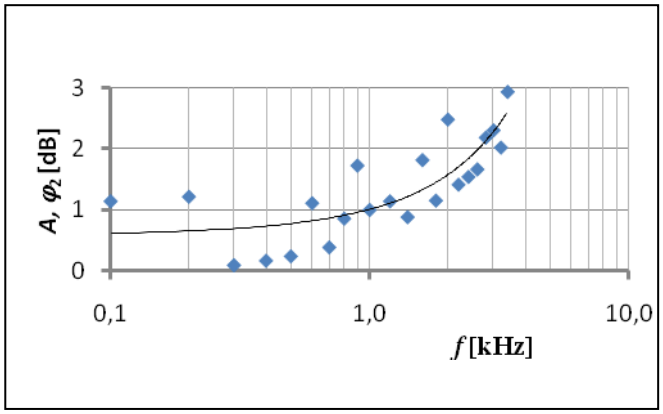


Figure 7. Relation between attenuation and frequency for Z_2 = 1040 Ω

Similarly, the $\varphi_2(f)$ function is defined by the following relation (see Figure 7):

$$\varphi_2(f) = 0{,}0676\, f_2^2 + 0{,}3565\, f_2 + 0{,}5856\,. \tag{3}$$

The K constant from equation (1) was determined based on measured and calculated values for f = 800 Hz and Z_2 = 1040 Ω

$$K = \frac{A}{\varphi_1(Z_2)\cdot\varphi_2(f)} = \frac{0{,}862}{0{,}541\cdot 0{,}914} = 1{,}744\,. \tag{4}$$

To determine attenuation, the regression model uses the following equation:

$$A = 1{,}744\cdot\left(7\cdot 10^{-7} Z_2^2 - 0{,}0016 Z_2 + 1{,}448\right)\cdot \cdot\left(0{,}0676\, f^2 + 0{,}3565\, f + 0{,}05856\right)\,. \tag{5}$$

IV. Analytical Mathematical Model

The analytical mathematical model is based on known secondary parameters of the wiring and the determination of attenuation using the following equation:

$$A = 10\cdot\log\frac{P_1}{P_2} \qquad [\mathrm{dB};\mathrm{W}]\,, \tag{6}$$

where P_1 is input performance and P_2 output performance. If

$$P_2 = \frac{U_2^2}{Z_2} \qquad [\mathrm{W};\mathrm{V},\Omega] \tag{7}$$

and

$$P_1 = \frac{U_1^2}{Z_1} \qquad [\mathrm{W};\mathrm{V},\Omega]\,, \tag{8}$$

then the following equation is valid for attenuation:

$$A = 10\cdot\log\frac{U_1^2}{Z_1}\cdot\frac{Z_2}{U_2^2} = \\ = 20\cdot\log\frac{U_1}{U_2} + 10\cdot\log\frac{Z_2}{Z_1} \qquad [\mathrm{dB};\mathrm{V},\Omega] \tag{9}$$

The input impedance is determined as follows:

$$\overline{Z_1} = \overline{Z_0}\cdot\frac{\overline{Z_2}\cdot\cosh\left(\overline{\gamma}\cdot l\right) + \overline{Z_0}\cdot\sinh\left(\overline{\gamma}\cdot l\right)}{\overline{Z_2}\cdot\sinh\left(\overline{\gamma}\cdot l\right) + \overline{Z_0}\cdot\cosh\left(\overline{\gamma}\cdot l\right)}\,. \tag{10}$$

To determine output voltage, the following equation can be used:

$$\overline{U_2} = \frac{\overline{U_1}}{\cosh\left(\overline{\gamma}\cdot l\right) + \frac{\overline{Z_0}}{\overline{Z_2}}\cdot\sinh\left(\overline{\gamma}\cdot l\right)}\,. \tag{11}$$

The exact wiring solution should respect the balanced placement of parameters over the complete wiring. Momentary current and voltage values are a function of both time and the place in which the values are being determined [4], [7].

In a harmonic regime, the behaviour of a long line is described by differential equations with effective voltage and current values. Both voltage and current are mere functions of the coordinates, they are constant in time.

Accordingly, equations (10) and (11) have been derived from the known cascade equations of the line.

$$\overline{U_1} = \overline{U_2}\cdot\cosh\left(\overline{\gamma}\cdot l\right) + \overline{Z_0}\cdot\overline{I_2}\cdot\sinh\left(\overline{\gamma}\cdot l\right) \\ \overline{I_1} = \frac{\overline{U_2}}{\overline{Z_0}}\cdot\sinh\left(\overline{\gamma}\cdot l\right) + \overline{I_2}\cdot\cosh\left(\overline{\gamma}\cdot l\right)\,. \tag{12}$$

V. Conclusion

Figures 8 and 9 illustrate the comparison of the tightness and appropriateness of the regression mathematical model is compared with an analytical model. The figures clearly display a similar progress in both models. The tightness of regression substitutes is expressed by a regression

coefficient. It equals R_2 = 0.56 for Figure 8 and R_2 = 0.65 for Figure 9. Deviations from the attenuation progress up to the load impedance of approximately 500 Ω for frequencies up to 1 kHz are due to equipment used. Its precision deviation is approximately 0.5 %.

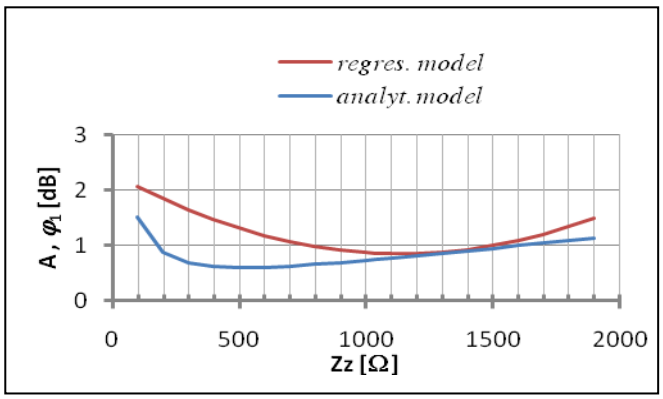


Figure 8. Relation between attenuation and load impedance for f = 800 Hz

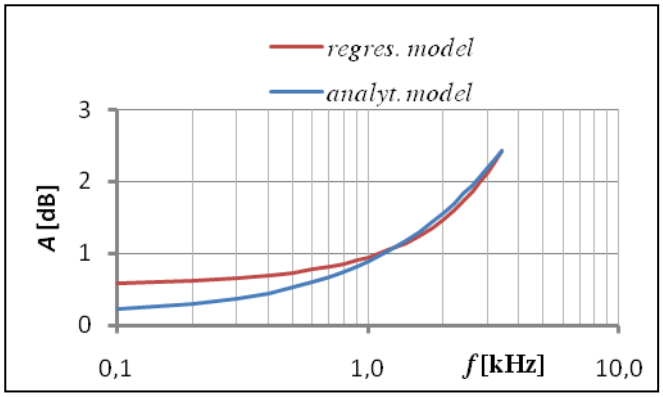


Figure 9. Relation between attenuation and frequency for Z_2 = 1040 Ω

The benefit of the regression mathematical model is mainly its applicability in practice. Based on the measured load impedance of a subscriber line, equation (5) enables to determine the relation between attenuation and frequency for transfers without any interference. The subsequent measurements of the relation between attenuation and frequency for a real line with interferences and their comparison with the determined course enable identifying the interfering frequencies. By analysing the frequencies, the sources of interference can be identified.

Acknowledgment

This research and results thereof were funded from the 7th Framework Programme of the European Communities (FP7/2007-2013) based on the grant agreement no. 218086, the INDECT project.

References

[1] BRIŠ R., LITSCHMANNOVÁ M., *Statistika I. pro kombinované a distanční studium*, Elektronické skriptum VŠB TU Ostrava, 2004.

[2] PETŘÍKOVÁ, I: Analýza kvality přenosu signálů na účastnickém vedení. Referát k doktorské zkoušce. (2009).

[3] PETŘÍKOVÁ, Iva. Analýza signálů na účastnickém vedení. *EE - Časopis pro elektrotechniku a energetiku : Zborník ku konferencii Elektrotechnika, Informatika a Telekomunikácie 2010*. Október 2010, 16, Mimoriadne číslo, s. 102-105. Dostupný také z WWW: <www.casopisee.sk>. ISSN 1335-2547. [článek]

[4] SCHLITTER, Miloš. *Telekomunikační vedení*. Praha : ČVUT, 1995. 258 s.

[5] VODRÁŽKA, Jiří. Přenos vysokými rychlostmi na symetrických párech. Praha : ČVUT FEL, 2000. Disertační práce na Katedře telekomunikační techniky.

[6] LAFATA, Pavel; VODRÁŽKA, Jiří. Modeling of Transmission Functions and Crosstalk in Metallic Cables for Implementation of MIMO Concept. *Radioengineering* [online]. December 2009, 18, 4, [cit. 2011-03-29]. Dostupný z WWW: <http://www.radioeng.cz>. ISSN 1210-2512. [e-článek]

[7] VODRÁŽKA, Jiří. Multi-carrier modulation and MIMO princinple application on subscriber lines. *Radioengineering* [online]. December, 16, 4, [cit. 2011-03-29]. Dostupný z WWW: <http://www.radioeng.cz>. ISSN 1210-2512. [e-článek]

Analysis of the dynamic range of electronic filter input signals

Zdenek Tesar
Department of Telecommunications
VSB–Technical University of Ostrava, Faculty of Electrical Engineering and Computer Science
17. listopadu 15, 708 33 Ostrava – Poruba, Czech Republic
zdenek.tesar@vsb.cz

***Abstract* – The paper deals with the impact of static and dynamic nonlinearities inherent to modern active components such as VFA, CFA, OTA and CCII on the dynamic range of the electronic filter input signal. Frequently, this impact is disregarded when analysing and synthesising filters although it is very important from the point of the view of active filter implementation. This is why the paper defines criteria for a comprehensive evaluation of filter circuit solutions to be used for the purpose of filter analysis and synthesis. Criteria have been defined while taking into account the maximum dynamic range of the input signal, and the paper also describes their application and relating recommendations for filter design.**

Key words: electronic filter, dynamic range of input signal, static and dynamic nonlinearity, active components VFA, CFA, OTA and CCII+.

INTRODUCTION

To design electronic filters, a number of circuit solutions can nowadays be applied. These solutions involve classic or modern active components that can work in voltage, mixed or current mode. Active frequency filters have a number of advantages but their disadvantage is the limited frequency range and the range of input signal. Once the maximum input signal range is exceeded, the filter is no longer a linear circuit and its designed parameters are degraded. This degradation occurs whenever a certain filter component stops working in the linear range of its characteristics. In order to ensure the maximum level of the input signal (maximum input dynamic range) of filters operating in voltage, mixed as well as current mode, the optimal size of the transfers from input into the critical bundles and branches needs to be secured.

I. NONLINEARITY OF ACTIVE COMPONENTS

To be able to classify a filter as linear, all passive and active components have to operate in the linear range and their nonlinear parameters are not applied. In general, any electronic circuit and filter can be subdivided into a passive and active part. We will consider the passive part as linear (almost always the case) for the analysed levels of signal. The filter nonlinearity will be derived from the nonlinearity of active components. For the analysis of input signal dynamic range, we will define the nonlinearities of commercially available active components. These include operational amplifiers with voltage feedback (VFA), transconductance amplifiers (OFA), current feedback amplifiers (CFA) and conventional current conveyor (CCII+).

Nonlinearities of active components can be subdivided into static (limits of signal amplitude) and dynamic (limit of the signal change speed), [1].

As for the standard VFA operational amplifier, static nonlinear parameters are a function of the input voltage and the output current (current limitation) – see Figure 1. For a harmonic signal, the dynamic nonlinear parameter – slew rate – is a function of the marginal output frequency f_P and the output voltage amplitude U_0 by

$$SR = \frac{\Delta U_0}{\Delta t} = 2\pi f_P U_0 \qquad (1)$$

$U_{0\,max}$ is the maximum output voltage in the linear area.

$I_{0\,max}$ is the maximum output current in the linear area.

$U_{cm\ max}$ is the maximum concurrent voltage in the linear area.

SR_{max} is the maximum slew rate of the input voltage in the linear area.

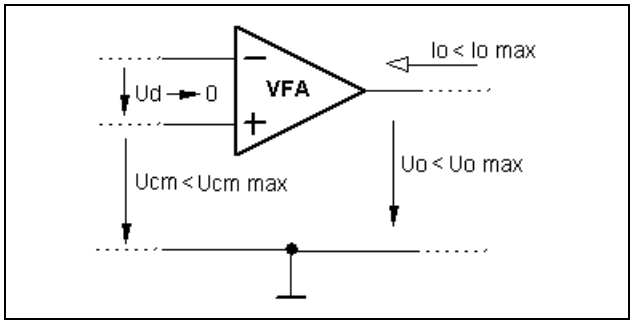


Figure 1. Static nonlinear parameters of VFA operational amplifier

Nonlinearities of the OTA transconductance amplifier are defined as described on Figure 2.

$U_{iP\ max}$ is the maximum noninverse input voltage in the linear area.

$U_{iN\ max}$ is the maximum inverse input voltage in the linear area.

$U_{0\,max}$ is the maximum voltage at the current output in the linear area.

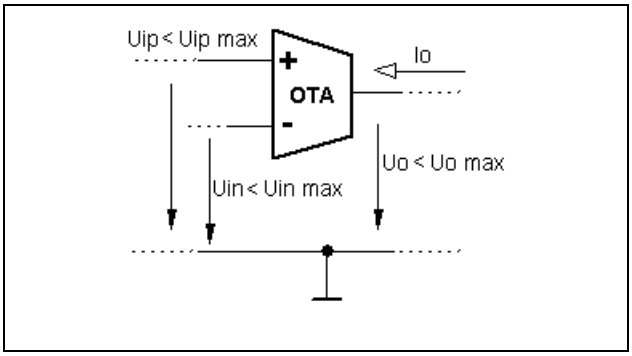


Figure 2. Static nonlinear parameters of the OTA transconductance amplifier

Again, these static nonlinear parameters are a function of the supply voltage. The voltage on the current output is a function of the total load impedance. The importance of the dynamic nonlinear parameter *SR* is similar as with the VFA amplifier. Figure 3 and 4 show the importance of static nonlinear parameters for CFA amplifier and CCII+ conveyor respectively. The parameters are the same for both types of circuits. The importance of dynamic nonlinear parameter *SR* is similar as with VFA amplifier – see equation (1). Given the principles of their activity, SR is theoretically unlimited in such circuits. In practice, it is always limited but can nevertheless reach high values.

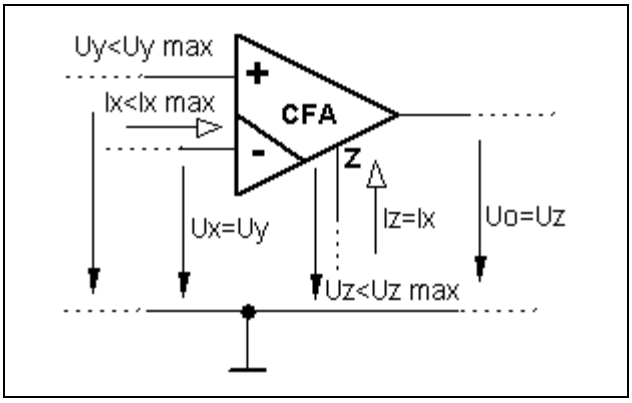


Figure 3. Static nonlinear parameters of CFA amplifier with current feedback

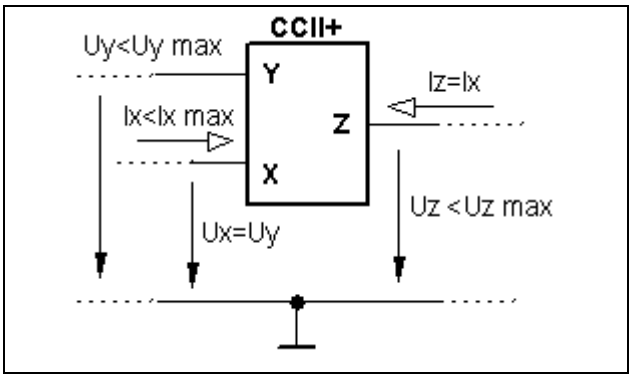


Figure 4. Static nonlinear parameters of CCII+ current conveyor

$U_{Y\ max}$ is the maximum input voltage on a non-inverting (high-impedance) input in the linear area.

$I_{X\ max}$ is the maximum input current on an inverting (low-impedance) input in the linear area.

$U_{Z\ max}$ is the maximum voltage at the current output "z" in the linear area.

Again, static nonlinear parameters are a function of the input voltage. The "z" output voltage is determined based the total impedance attached to a particular contact.

II. CRITERIA DETERMINING VOLUME OF OUTPUT SIGNAL

The basic criteria for comparing and assessing individual filter circuit solutions from the point of view of the input signal dynamic range are determined using the analysis of the operation network's general structure with VFA amplifiers as shown on Figure 5 and its modification for other mentioned active components OTA, CFA and CCII+ (Figure 6).

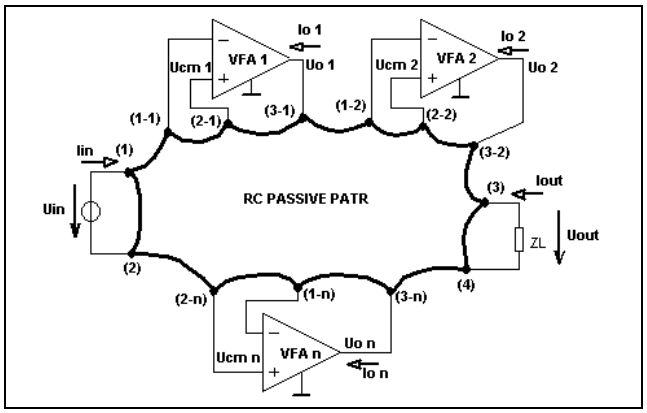


Figure 5. Operation network with VFA

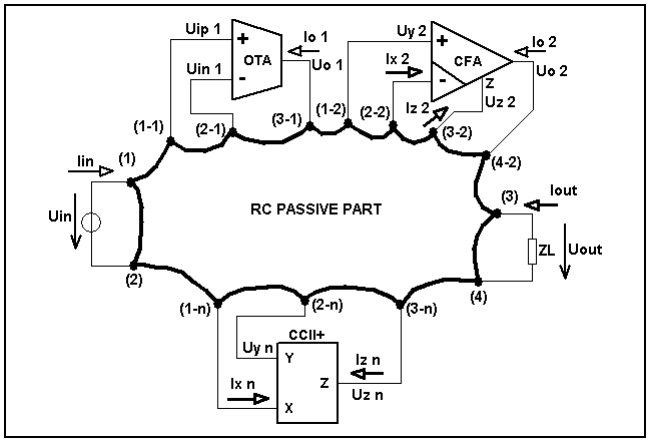


Figure 6. Operation network with OTA, CFA and CCII+ circuits

Based on the analysis of a particular filter configuration, we define critical bundles **N** and branches **V** corresponding with active components' input and output for key nonlinearities of active components, as shown on Figure 1 – Figure 4.

Criterion no. 1

We determine the input voltage transfer frequency relations at filter's critical bundles:

$$F_{VN}(f) = \frac{U_N}{U_{IN}}. \qquad (2)$$

We determine the maximum voltage transfer in a given frequency band:

$$F_{VN\max} = \max\{F_{VN}(f)\} \tag{3}$$

And for

$$U_{N\max} = \min\begin{Bmatrix} U_{cm\max}, U_{iP\max}, U_{iN\max}, \\ U_{0\max}, U_{Y\max}, U_{Z\max} \end{Bmatrix} \tag{4}$$

We define the maximum input voltage:

$${}^{F_V}U_{IN\max} < \frac{U_{N\max}}{F_{N\max}}. \tag{5}$$

Criterion no. 2

We determine the transfer admittance frequency relation from input into filter critical branches:

$$Y_{TV}(f) = \frac{I_V}{U_{IN}}. \tag{6}$$

We determine the maximum transfer admittance in a given frequency band:

$$Y_{TV\max} = \max\{Y_{TV}(f)\} \tag{7}$$

And for

$$I_{V\max} = \min\{I_{0\max}, I_{X\max}\} \tag{8}$$

We determine the maximum input voltage:

$${}^{Y_T}U_{IN\max} < \frac{I_{V\max}}{Y_{TV\max}}. \tag{9}$$

Criterion no. 3

We determine frequency relations of input current transfers into filter's critical branches:

$$F_{CV}(f) = \frac{I_V}{I_{IN}} \tag{10}$$

We determine the maximum current transfer in a given frequency band:

$$F_{CV\max} = \max\{F_{CV}(f)\} \tag{11}$$

And for

$$I_{V\max} = \min\{I_{0\max}, I_{X\max}\} \tag{12}$$

We determine the maximum input voltage:

$${}^{F_C}I_{IN\max} < \frac{I_{V\max}}{F_{CV\max}}. \tag{13}$$

Criterion no. 4

We determine the frequency relation of input transfer impedance into filter's critical bundles.

$$Z_{TN}(f) = \frac{U_N}{I_{IN}}. \tag{14}$$

We determine the maximum transfer impendence for a given frequency band

$$Z_{TN\max} = \max\{Z_{TN}(f)\} \tag{15}$$

And for

$$U_{N\max} = \min\begin{Bmatrix} U_{cm\max}, U_{iP\max}, U_{iN\max}, \\ U_{0\max}, U_{Y\max}, U_{Z\max} \end{Bmatrix} \tag{16}$$

We determine the maximum input current

$${}^{Z_T}I_{IN\max} < \frac{U_{N\max}}{Z_{TN\max}}. \tag{17}$$

The maximum dynamic range of input signal of an analysed filter for input voltage is determined as

$$U_{IN\max} = \min\{{}^{F_V}U_{IN\max}, {}^{Y_T}U_{IN\max}\} \tag{18}$$

And for the input current as

$$I_{IN\max} = \min\{{}^{F_C}I_{IN\max}, {}^{Z_T}I_{IN\max}\}. \tag{19}$$

Criterion no. 5

We establish f_{max} frequencies at which the volume of voltage transfer and input transfer impedance reach their maximum into active components' voltage output. We also test whether the following condition is fulfilled for a particular SR of the active component

$$SR > 2\pi f_{\max} F_{VN\max}(f_{\max}) U_{IN\max} \tag{20}$$

or

$$SR > 2\pi f_{\max} Z_{TN\max}(f_{\max}) I_{IN\max}. \tag{21}$$

Criteria 1 and 2 have been designed for filters with voltage input operating in the voltage or mixed mode. Criteria 3 and 4 have been designed for filters with current input that operate in the current or mixed mode. These criteria take into account static nonlinearities of active components. Criterion 5 takes into account dynamic nonlinearities. Filter circuit solution and implemented active components determine which criterion is the key to maximise the dynamic range.

When designing the filter, we implement such adjustments of circuit solution or components enabling to minimise all critical transfers.

III. VERIFICATION BY EXPERIMENT

The assessment of filter configuration using the above described criteria needs to be carried out in particular when designing filters with a higher number of active components; filters with high quality factor; high-order filters implemented using cascade and non-cascade synthesis as well as for retuned filters. We have applied the conclusions described in this paper in designing a second-order bandpass with CCII+ components (see Reference [2]). Its parameters are as follows: $F_{PP}(f_0)$ = 1, Q = 10, f_0 = 100kHz. Figure 7 shows the configuration. Critical terminals (1) – input, (2) – output, and (3) were defined for the given bandpass configuration. Table 1 provides an overview of design relations and transfers $F_{21}=U_2/U_1$ a $F_{31}=U_3/U_1$; Figure 8 shows transfer frequency characteristics.

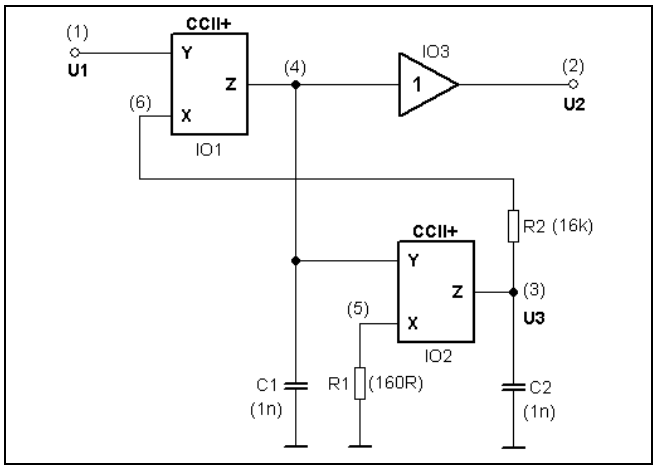


Figure 7. Bandpass PP1 Q=10

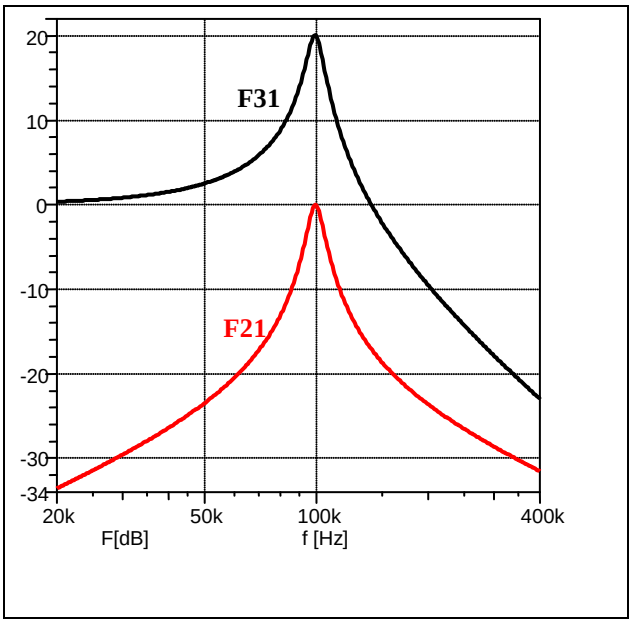


Figure 8. Frequency characteristics PP1 Q=10

TABLE I. DESIGN RELATIONS AND TRANSFERS FOR PP1

$$F_{21}(f_0)=\frac{C_2}{C_1} \quad (22) \qquad Q=\sqrt{\frac{R_2C_2}{R_1C_1}} \quad (23)$$

$$f_0=\frac{1}{2\pi\sqrt{R_1R_2C_1C_2}} \quad (24)$$

$$F_{31\max}\approx\sqrt{Q^2+1} \quad (25)$$

$$F_{21}=\frac{U_2}{U_1}=\frac{p\frac{1}{R_2C_1}}{p^2+p\frac{1}{R_2C_2}+\frac{1}{R_1R_2C_1C_2}} \quad (26)$$

$$F_{31}=\frac{U_3}{U_1}=\frac{p\frac{1}{R_2C_2}+\frac{1}{R_1R_2C_2C_1}}{p^2+p\frac{1}{R_2C_2}+\frac{1}{R_1R_2C_1C_2}} \quad (27)$$

F_{21} and F_{31} transfer frequency curves show that the maximum value of F_{21} transfer equals one in compliance with the assignment. As for F_{31} transfer, its maximum value as determined by equation (25) is ten times higher. It increases with the increasing quality factor Q of the pass, i.e. the maximum admissible amplitude of the input signal decreases. To put it simply, where the amplitude of the bandpass input voltage isU_{1PP} = 1V, the output voltage is to have the same amplitude U_{2PP} = 1V. The voltage amplitude of terminal (3) reaches U_{3PP} = 10V. Where the supply voltage is for example ±5V, the IO2 conveyor is not going to operate in the linear area. Accordingly, the described bandpass configuration is to yield small input voltage amplitude for higher quality factors. It is not possible to extend the input signal dynamic range in this configuration by modifying the values of passive components (see design relations in Table 1). Using an attenuator at filter input and the subsequent increase of the output signal is not the optimum solution either. The solution is to apply sub-critical voltage positive feedback (bootstrap) represented by R_A and R_B resistors that increases the filter quality agent's value, see References [3] and [4]. Figure 9 shows the modified PP2 configuration with the same parameters but a significantly higher dynamic range (approx. 10 times). Table 2 shows design relations and transfer relations $F_{21}=U_2/U_1$, $F_{31}=U_3/U_1$ a $F_{71}=U_7/U_1$. Figure 10 show their frequency characteristics.

A comparison of equations (32), (36) and (37) shows that the input signal amplitude is most influenced by the maximum transfer volume $F_{31\beta}$ which is a function of the bandpass quality agent without feedback (see relation (36)). In our design, we opted for low bandpass quality agent without feedback Q=0.5. This is how we minimised the $F_{31\beta}$ transfer. The required value of the Q_β quality agent is set using sub-critical positive feedback, R_A and R_B resistors (see relations (30) and (31)).

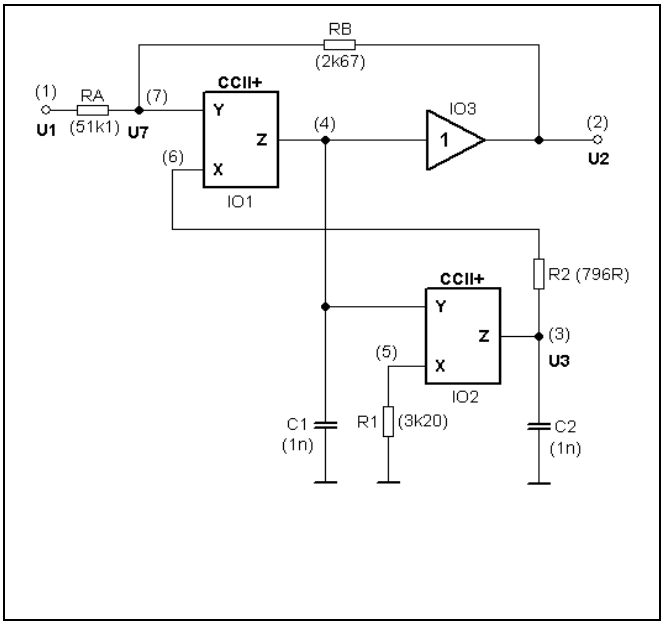


Figure 9. Modified bandpass PP2 Q=10

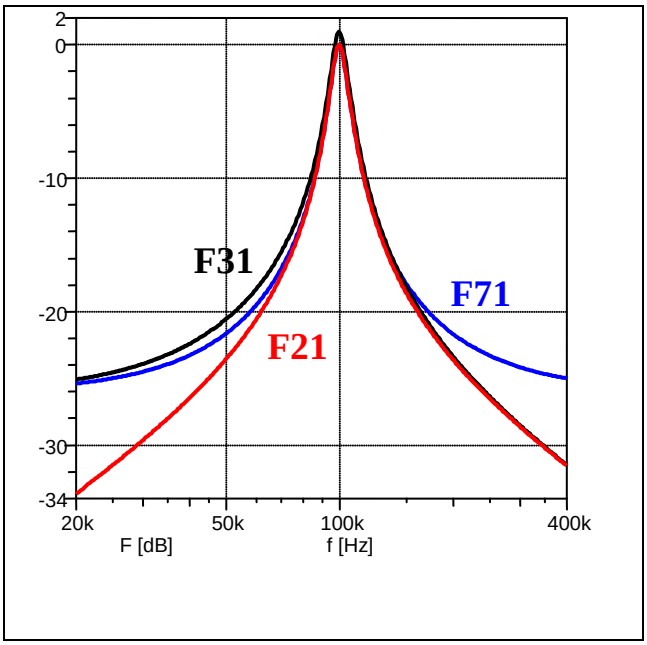


Figure 10. Frequency characteristics PP2 Q=10

The stability of the filter needs to be paid attention to. The following equation should be valid (38).

$$1-\frac{C_2}{C_1}\beta>0 \qquad (38)$$

Figure 10 displays the frequency characteristics of individual transfers. It is clear that maximum transfer volumes approach the required value one (0 dB) and that the input signal dynamic range is used to its maximum.

TABLE II. DESIGN RELATIONS AND TRANSFERS FOR PP2

$$f_0=\frac{1}{2\pi\sqrt{R_1R_2C_1C_2}} \quad (28) \qquad \beta=\frac{R_A}{R_A+R_B} \quad (29)$$

$$1-\beta=\frac{R_B}{R_A+R_B} \quad (30)$$

$$Q_\beta=\frac{1}{1-\frac{C_2}{C_1}\beta}\sqrt{\frac{R_2C_2}{R_1C_1}}=\frac{1}{1-\frac{C_2}{C_1}\beta}Q \quad (31)$$

$$F_{21\beta}(f_0)=\frac{C_2}{C_1}\frac{1-\beta}{1-\frac{C_2}{C_1}\beta}=F_{21}(f_0)\frac{1-\beta}{1-\frac{C_2}{C_1}\beta} \quad (32)$$

$$F_{21\beta}=\frac{p\frac{1}{R_2C_1}(1-\beta)}{p^2+p\frac{1}{R_2C_2}\left(1-\frac{C_2}{C_1}\beta\right)+\frac{1}{R_1R_2C_1C_2}} \quad (33)$$

$$F_{31\beta}=\frac{\left(p\frac{1}{R_2C_2}+\frac{1}{R_1R_2C_1C_2}\right)(1-\beta)}{p^2+p\frac{1}{R_2C_2}\left(1-\frac{C_2}{C_1}\beta\right)+\frac{1}{R_1R_2C_1C_2}} \quad (34)$$

$$F_{71\beta}=\frac{\left(p^2+p\frac{1}{R_2C_2}+\frac{1}{R_1R_2C_1C_2}\right)(1-\beta)}{p^2+p\frac{1}{R_2C_2}\left(1-\frac{C_2}{C_1}\beta\right)+\frac{1}{R_1R_2C_1C_2}} \quad (35)$$

$$F_{31\beta\max}\approx\frac{1-\beta}{1-\frac{C_2}{C_1}\beta}\sqrt{Q^2+1} \quad (36)$$

$$F_{71\beta\max}\approx\frac{1-\beta}{1-\frac{C_2}{C_1}\beta} \quad (37)$$

IV. CONCLUSION

The maximum admissible dynamic range of the input signal is very important from the point of view of the proper filter functioning. Sometimes this is not considered sufficiently when designing the filters. This paper defines basic criteria to assess filter circuit solutions with commercially available active components (VFA, CFA, OTA and CCII+) and taken into consideration the dynamic range of the input signal. The bandpass solution illustrates the methodology for dynamic range optimisation by adjusting the filter configuration (implementing sub-critical positive feedback). Another optimisation method is choosing the optimum values of filter passive components – for more details see Reference [5].

ACKNOWLEDGEMENT

This research and results thereof were funded from the 7th Framework Programme of the European Communities (FP7/2007-2013) based on the grant agreement no. 218086, the INDECT project.

REFERENCES

[1] DOSTÁL, J.: Operační zesilovače. Praha: SNTL, 1981.

[2] FABRE, A.: High Input Impedance Insensitive Second-Order Filters Implemented from Current Conveyors. IEEE Trans. CAS-I, Vol. 41 No. 12, 1994, 918-921

[3] DOSTÁL, T.: Elektrické filtry. Brno: skriptum VUT, 2007.

[4] PUNČOCHÁŘ, J.: Operační zesilovače v elektronice. Praha: BEN, 1996. ISBN 80-901984-3-0.

[5] VALKO, R., LATTENBERG, I.: Dynamická optimalizace rozsahu vstupního signálu filtrů s transimpedančními zesilovači. www.elektrorevue.cz, 2005/45

Author Index