by Anhalt University of Applied Sciences and Perm National Research Polytechnic University

Proceedings of 1st International Conference on Applied Innovations in IT

Eduard Siemens (editor in chief) et al.

ISBN 978-3-86011-054-6 (Print)

EDITION
Hochschule Anhalt

Proceedings of 1st International Conference on Applied Innovations in IT

Eduard Siemens (editor in chief) et al.

This volume contains publications of the 1st International Conference on Applied Innovations in IT (ICAIIT), which took place in Koethen March 25th 2013. The conference is devoted to problems of applied research in the fields of mechanical and economical engineering, auotmation and communications as well as of data mining. The research results can be of interest for researchers and development engineers, who deal with theoretical base and the application of the knowledge in the respective areas.

Print: Anhalt University of Applied Sciences, Dessau 2013

Proceedings of 1st International Conference on Applied Innovations in IT, 2013.

editors: Prof. Dr. Eduard Siemens* (editor in chief),
Prof. Dr. Alexander Kostygov**, Dr. Bernd Krause*, Dr. Leonid Mylnikov**
(*Anhalt University of Applied Sciences, ** Perm National Research Polytechnic University)

ISBN 978-3-86011-054-6 (Print)
ISBN 978-3-86011-055-3 (Online)

Content

Electron Beam Welding In-Process Control and Monitoring

Dmitriy Trushnikov
Perm National Research Polytechnic University – Mechanic und Technology department
Komsomolsky avenue 29, 614990 Perm, Russia
E-mail: trdimitr@yandex.ru

***Abstract*—This work presents the results of an investigation of processes in the melting zone during Electron Beam Welding (EBW) through analysis of the secondary current in the plasma.The studies show that the spectrum of the secondary emission signal during steel welding has a pronounced periodic component at a frequency of around 15–25 kHz. The signal contains quasi-periodic sharp peaks (impulses). These impulses have stochastically varying amplitude and follow each other in series, at random intervals between series. The impulses have a considerable current (up to 0.5 A). It was established that during electron-beam welding with the focal spot scanning these impulses follow each other almost periodically. It was shown that the probability of occurrence of these high-frequency perturbation increases with the concentration of energy in the interaction zone. The paper also presents hypotheses for the mechanism of the formation of the high-frequency oscillations in the secondary current signal in the plasma.**

***Keywords:* automation of welding processes; electron beam welding; focus control; focus spot scanning; weld formation monitoring**

I. Introduction

ELECTRON beam welding (EBW) is a fusion welding process often done in a vacuum. The process has a number of advantages: high power concentration in the electron beam, easy control of the energy flow into the metal, smaller heat-affected areas, equal strength of the weld joint and main metal, etc. These advantages allow the use of electron beam for welding reactive and nonferrous metals, high-tensile and heat-resistant alloys that are typically used in the production of critical products.

However, certain problems arise in EBW process, which are related to instability of weld-joint formation and difficulties in creating and controlling an optimal focus regime. A major limitation in controlling such a focus regime is the lack of understanding of the processes during EBW. The complex character and high speed of these processes make numerical modeling very difficult, forcing researchers to rely on experimental research methods.

The reported study here was partially supported by RFBR, research projects No.11-08-96016-a and No.13-08-00397A and by Ministry of Education of Perm district of the Russian Federation.

D. N. Trushnikov is with the Department of Applied Physics, Perm National Research Polytechnic University, Russia, Perm, Komsomolsky Av. 29, 614990 (e-mail: trdimitr@ yandex.ru).

The basic parameters of EBW are accelerating voltage, electron beam current, focusing-coil current, welding speed, operating gun-sample distance, vacuum level in the process chamber, etc. These parameters are chosen according to factors such as the operator's own experience, mathematical models [1, 2], or statistical analysis [3–4]. The most difficult parameter to identify and reproduce in EBW is the focusing position. The operator of an EBW needs to manually set the focus of the beam. The adjustment of the focusing-coil current is based on the subjective operator evaluation of luminosity brightness, emitted from the interaction area of the beam irradiating refractory target material, e.g. wolfram. When the luminosity brightness becomes maximal, the focusing mode is considered sharp [5].

The process of manual focus control is subjective and can lead to performance depreciation. Each operator interprets the luminosity brightness of the operational area differently and, therefore, the welding results could be not reproducible. Changing the focusing current by 1% may cause a 20–60% fluctuation of fusion depth. The focusing position also significantly influences the probability of various defects specific to EBW such as spiking, cavitations, medial cracks, etc. The difficulties in focusing control are aggravated by changes in the electronic and optical systems of an electronic beam gun due to cathode wear and tear or after planned maintenance.

In recent years, this problem has been partly solved by using a modified Faraday cup to control the electronic beam density distribution [6]. During circular scanning, the beam passes through a set of radial gaps in the wolfram disk. After the current passing through the gaps is measured, the density of the electronic-beam power, beam diameter, maximum specific power and other important metrics are calculated based on computer tomography algorithms. By controlling these basic parameters of the electronic beam, the parameters of the welding seams can be reproduced. It has also been reported that due to this measurement of beam characteristics, the concrete EBW technology can be migrated between various electron beam sets [7]. Some vacuum chambers do not support internal mounting of the required sensor (the modified Faraday cup). For welding operation modes, the focusing current should be adjusted based on experiments with various materials, thicknesses and types of electronic-beam guns. Moreover, the systems based on modified Faraday cup do not

executed the real-time control and adjustment of the focusing mode during welding. Real-time adjustments are important for large objects welding especially when the cathode electron emission and thus the electronic and optic adjustments of the gun are significant.

Therefore, the control, monitoring and analysis of the processes in welding bath during EBW requires analysis of the secondary signal parameters, such as secondary electron or ion emission, optical emission, X-rays, etc.

One of the specific processes caused by the impact of the dense electron beam to the metal during EBW is the formation of plasma in the operational area [8–9]. The parameters of the plasma are closely connected with the electron beam thermal effect on the metal being welded. In [10–11], plasma current parameters are suggested for electron beam focusing control.

All the above methods use extreme correlations between the secondary emissions and the focusing lens current. These correlations are characterized by the dead zones and two values of the focusing lens current that ensure similar signal parameters, but various derivatives. The application of focal spot scanning (modulation of the focusing lens current) was required in this case for the embodiment of the methods of operational control. All known research has been based on the extreme of a signal function, obtained while slowly changing the focusing current. In this case, the execution of the operational control requires low-frequency scanning of the focusing current, which negatively affects the quality of the welded joint. High-frequency scanning of the focal spot for the purpose of improving the quality of the welded joint is known [12]. However, its applicability to the operational control of beam focusing has not been investigated until now.

In recent years, control and monitoring of electron beam and laser beam welding has become more and more popular [13–16]. Laser technologies and EBW are based on similar principles and used concentrated energy beams. New research opportunity provided by modern signal processing is finding an increased interest by researchers

One of the well-known ways to increase the signal/noise ratio for physical object analysis is periodic exposure of the object with subsequent analysis of its response at a given exposure frequency [17]. In [10,11,18], it is reported that periodic changes of the electron beam parameters (beam current modulation, oscillation or focus spot scanning) causes normalization of the processes in the keyhole created by the electron beam, thus making the series of impulses passing at constant interval`s proportional to the waveform frequency. However, obtained results are not sufficiently reliable, so further investigation is required. For example, simultaneous recording the focusing lens current and of the secondary current signals collected by the plasma was not performed. Also, the principles for generating harmonics proportional to the scanning frequency were not revealed.

This article studies the behavior of the current collected from the plasma, generated in the operational area of the electron beam, when using EBW with focal spot scanning (modulation of the focusing lens current), based on the coherent accumulation method (cross-correlation analysis) [17, 19]. This method can be used to obtain not only amplitude ratios, but also the phase ones, as well as determining how the current signals in the plasma are synchronized with the focusing lens signals during EBW. These results can be useful for focal spot scanning - parameters selection methods and methods to control EBW against the parameters of the plasma current.

II. . Experimental procedure

A ring electrode collector was used to measure the secondary current from the plasma. The collector was located over the zone of welding. The collector has a positive potential of 50 V. The loading resistance was 50 Ω. The signal from the collector was registered by a data acquisition system and further processed by a computer. The sampling frequency in the experiments was in a range from 100 kHz to 1 MHz per channel.

During the experiments, samples of chrome-molybdenum steel (0.15 % carbon, 5 % chrome and around 1 % of molybdenum) and high-alloy chrome-nickel steel (up to 0.12 % carbon, 18 % chrome and up to 0.8% titanium) were welded. The accelerating voltage in all experiments was 60 kV. The welding power was in the range from 2 kW to 4 kW.

During the experiments, the welding power P, welding speed, focus degree ΔI_f($\Delta I_f = I_f - I_{f0}$ is the difference between the average focusing lens current of the welding mode and the focusing lens current of sharp focus), the frequency f and amplitude of the focal spot scanning A were varied.

The current in the focusing lens was changed under a linear law. The limits of scanning frequency were from 90 to 12000 Hz. The amplitude of these oscillations was in a range from 3 to 25 mA.

Transverse metallurgical sections of the weld were made from all the welded samples. The focus regime was determined by the transverse sizes of the penetration depth. The sharp focus regime corresponds to the maximum penetration depth.

III. Results and discussion

Fig. 1 shows a typical spectrum of the secondary current signal collected from plasma during welding of steel samples. It can be noted that there is a characteristic maximum in the signal at frequencies close to 15-20 kHz. Fig. 2a shows the secondary signal together with a signal of focusing lens. The processes in the keyhole become periodic with focal spot scanning s. Frequency perturbations in the secondary current start to periodically follow at multiples of the scanning frequency. More detailed consideration (Fig. 2b) shows that each perturbation represents a series of high-frequency quasi-periodic impulses. Their frequency 10–25 kHz is very stable and specific for different materials and welding regimes. The amplitude changes randomly. The spectrums and waveforms of the secondary current in a plasma during electron-beam welding are more fully described in [10,11, 18].

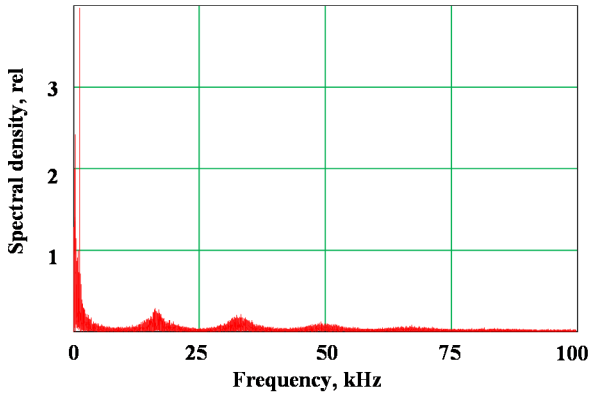


Fig. 1 A typical signal spectrum of the secondary current collected from the plasma during electron-beam welding with focus point oscillation (welding power: 2.5 kW, sharp focus regime (ΔI_f=0), scanning frequency: 561 Hz).

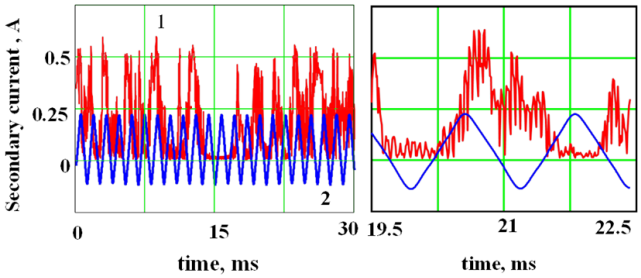


Fig. 2 Waveform of secondary current, collected from the plasma and the signal of the focus coil current during electron-beam welding with focus point oscillation. 1. *Data* (*t*): secondary current. 2.*Osc*(*t*): signal from the focusing lens current.

There is a hypothesis to explain the mechanism of the appearance of high-frequency oscillations in the secondary signal in the plasma. It deals with the assumption of the existence of explosive boiling in the keyhole [12, 20]. The rate of energy input in the interaction of the electron beam with the metal in the keyhole is much higher than the rate of heat removal through conduction. There is local overheating of the metal, followed by explosive boiling. The boiling metal vapor affects beam shielding, the beam is scattered by the metal vapor, and the power density is dramatically reduced. After the vapor evacuation from the keyhole, the beam power density is again above the critical and the process resumes. The frequencies predicted by this hypothesis are close to those observed experimentally (Fig. 1).

The described phenomena have important effect. They imply the extreme character of the probability of self-oscillation processes on the power density of the electron beam. The hypotheses do not disprove each other. Increasing the power density of the electron beam makes the difference between the velocities of the input energy in the metal and its removal even larger. This must increase the probability of local overheating of the metal.

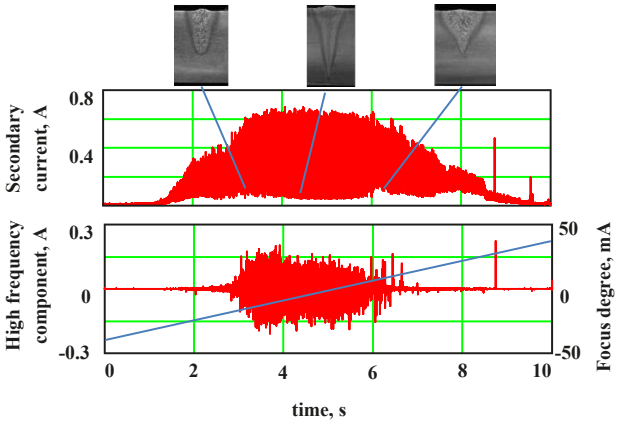


Fig. 3 The high-frequency component of the secondary current during the linear rise of the focusing lens current. (*P*=2 kW).

Fig. 3 shows the secondary current and the high-frequency component (*f*> 10 kHz) during the linear rise of the focusing lens current. The signal is appreciable in a certain range, accompanied by deep penetration. The high-frequency component appears in a narrow range. The dependence of the high-frequency component on the power density (the focus degree) is used to construct operational control methods.

In the given work, research into the secondary signal was conducted using coherent accumulation, which is an enhancement of coherent detection, and is widely applied to tracking an electronic beam on a seam, but it has been applied to research processes in the keyhole and welding control only recently. In this research, the high-frequency range 15–20 kHz was studied.

The coherent accumulation method is illustrated in Fig. 4. The small-width square-wave signal is formed from the signal from the current of the focusing lens (*Osc*(*t*)) - a basic signal *g*(*t*). The basic signal $g(t+\tau)$ is shifted relative to the initial signal *Osc*(*t*) for a set time τ (Fig. 4).

The signal of the secondary current, collected from the plasma $I_c(t)$ is processed by a digital or analog high-pass filter with a cutoff frequency around 10 kHz. The selected signal of the high-frequency component (*Data*(*t*)) is rectified and then multiplied by the basic signal, $g(t+\tau)$. The result is integrated over time. As a result, we have the function $S(\tau)$

$$S(\tau) = \int_0^{t_0} g(t+\tau) \cdot \left| Data(t) \right| dt ,$$

where t_0 is the sampling time. This function $S(\tau)$ expresses the average amplitude of the high-frequency secondary signal for each value of the shift kV. The welding power was in the range from 2 kW to 4 kW.

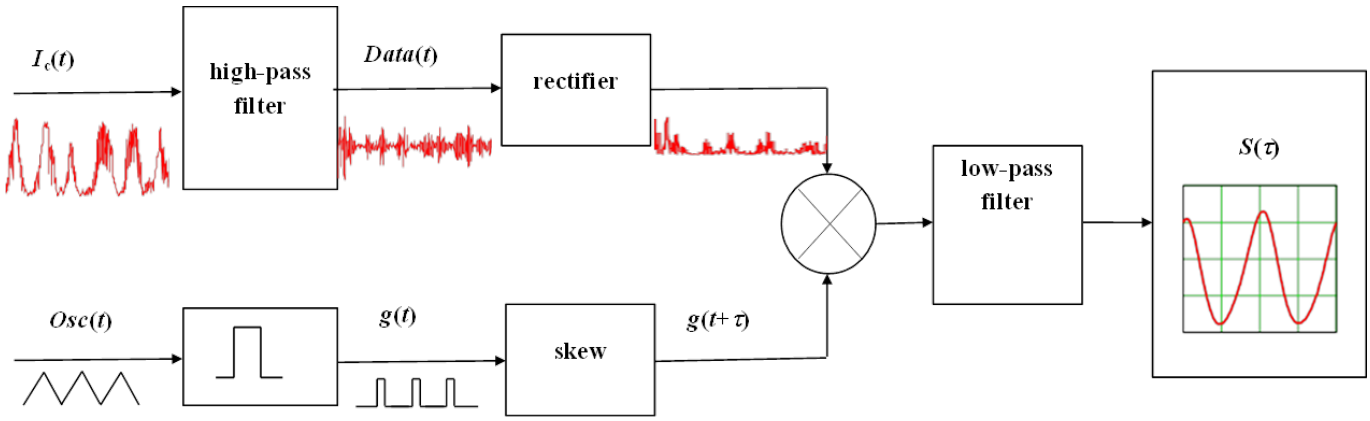


Fig. 3 Coherent accumulation method

In other words, for each value of focusing lens current, there is an average value of the amplitude of the high-frequency oscillations of the secondary signal.

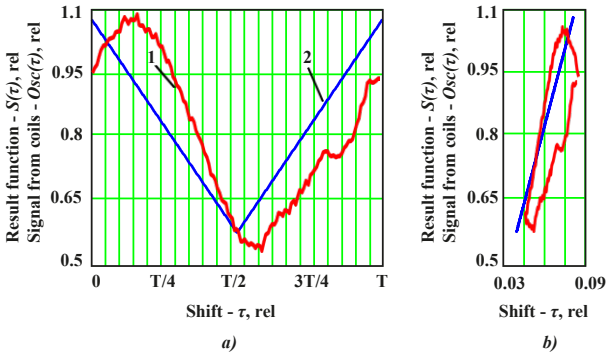


Fig. 5 1- Function $S(\tau)$, obtained using the coherent accumulation method on τ, is the result of secondary processing of the high-frequency component signal. 2- $Osc(\tau)$ is the record of the focusing lens current (P=2.5 kW, underfocus regime (ΔI_f= -10 mA), oscillation frequency f=966 Hz).

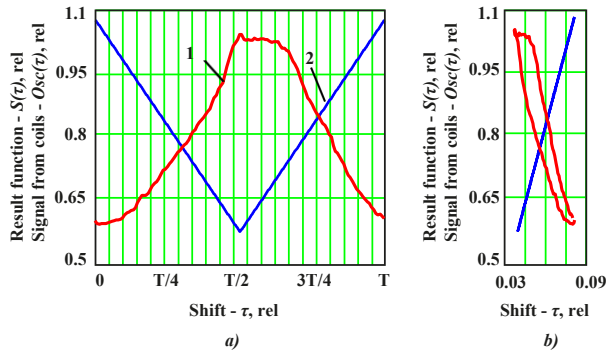


Fig. 6 1- Function $S(\tau)$, obtained using the coherent accumulation method on τ, is the result of secondary processing of the high-frequency component signal. 2- $Osc(\tau)$ is the record of the focusing lens current (P=2.5 kW, overfocus regime (ΔI_f= +17 mA), oscillation frequency f=966 Hz).

Fig. 5a shows the results of processing the secondary current signal using coherent accumulation with focal spot scanning . The underfocus regime was used. The frequency was 966 Hz and the amplitude of the focusing lens current oscillations was 7 mA. It is possible to present this function in phase space. For this purpose, on a horizontal axis we postpone the current of the focusing lens (Fig. 5b).The characteristic lag of the high-frequency component signal relative to the deflection coil current signal may be noted. A similar phenomenon has been observed in [21] and may be explained by thermal effects in the melting zone.

Fig. 6 shows the results when the beam is over-focused. The change in sign of the correlation coefficient when the beam is focused is of major interest. When the beam is under-focused the coefficient's sign is positive. As the focusing current is increased, the coefficient's magnitude decreases monotonically, becoming zero in the region of sharp focus. A similar phenomenon has been observed in the entire range of investigated conditions. The total number of observations in the multi-factor experiment was 107.

The change in the sign of the correlation coefficient during a change in the focusing current is highly significant. The existence of an extreme in the amplitudes of the high-frequency oscillations of the secondary current in the plasma as a function of the focusing current as it is slowly changed may be explained by the existence of an extreme in the welding parameters (weld penetration, width of the melting zone, etc.). In the experiments described, the frequency of change of the focusing current was several orders of magnitude larger than the frequencies characterizing the geometry of the melting zone. The results obtained confirm the hypothesis that the probability of the occurrence of high-frequency oscillations as well as the amplitude of those oscillations in the secondary waveform, increase monotonically given an increase in the concentration of energy in the area of interaction between the electron beam and the metal in the melting zone.

The function described makes it possible to detect focus conditions and directly control them in the process of EBW. To do this while welding with an oscillating beam, the electron gun's focusing lens current is modulated by frequency f. The current of the charged-particle collector installed over the electron beam weld zone is recorded during the welding process.

In this case the focusing current can be written as:

$$I_{\text{f}}(t) = \langle I_{\text{f}} \rangle \cdot (1 + \xi \cdot \cos \omega \cdot t)], \quad (1)$$

where $\langle I_{\text{f}} \rangle$ is the current average value of the focusing current; ξ is the modulation depth of the focusing lens current, $\omega = 2\pi f$ is the focus's scanning frequency (modulation of the focusing system's current).

The waveform of the secondary current in the plasma, which was captured by the electron collector, is subjected to high-frequency filtering and then rectification. In the process, the amplitude of the high-frequencies of the current captured by the electron collector will change according to the following law:

$$I_{\text{км}} = \varphi[I_{\text{fo}} - I_f(t)],$$

where $\varphi[I_{\text{fo}} - I_{\text{f}}(t)]$ – is the function described above which expresses the dependence of the amplitude of the high-frequency oscillations of the secondary current on the level of focus ($\Delta I_{\text{f}} = \langle I_{\text{f}} \rangle - I_{\text{fo}}$).

As shown above, in the absence of modulation of the focusing lens current ($\xi = 0$) the function $\varphi[I_{\text{fo}} - I_{\text{f}}(t)]$ has an extreme(maximum) when $I_{\text{f}} = I_{\text{fo}}$, and the vicinity of this point may be roughly approximated by the quadratic function

$$I_{\text{км}} = \varphi(I_{\text{fo}} - I_{\text{f}}) = a - k \cdot \Delta I_f{}^2 = a - k \cdot (I_{\text{fo}} - I_{\text{f}})^2, \quad (2)$$

where a, k are certain coefficients.

Inserting (1) into (2)

$$I_{\text{км}} = a - k \cdot (\Delta I_{\text{f}}{}^2 - 2 \cdot I_{\text{f}} \cdot \xi \cdot \Delta I_{\text{f}} \cdot \cos \omega \cdot t + I_{\text{f}}^2 \cdot \xi^2 \cdot \cos^2 \omega \cdot t) \quad (3)$$

If the waveform of (3), which expresses the amplitude of the high-frequency oscillations of the secondary current are synchronously demodulated at frequency ω then, after synchronous demodulation and low-frequency filtering in order to remove modulation oscillations, we obtain a waveform proportional to the expression:

$$I_{k\omega} = I_{\text{f}} \cdot \xi \cdot (I_{\text{f}} - I_{\text{fo}}) \quad (4)$$

This waveform makes it possible with high accuracy to determine and adjust the focusing current of an electron beam providing either the maximum weld penetration of the metal (a sharply focused beam) or a penetration value close to the maximum (over-focused or under-focused beams).

IV. Conclusion

1. In this paper, the ability to study processes in the penetration channel (keyhole) during electron-beam welding was demonstrated. A combination of an application of focus beam oscillation and an analysis of the instabilities of the secondary current, collected from the plasma over the welded samples was used for this purpose.

2. The experimentally obtained secondary current signal, collected from plasma during electron-beam welding with electron beam oscillation contains a series of high-frequency perturbations, which follow each other at certain frequencies that are multiples of the deflection scanning frequency.

3. It was shown that the probability of occurrence of these high-frequency perturbations increases with the concentration of energy in the interaction zone. Hypotheses for the emergence of high-frequency oscillating processes in the "beam-keyhole-plasma" structure were considered. The local overheating of the metal is possible during electron-beam welding with a subsequent vapor explosion and defocusing of the latter, when the vapor density reaches a critical value.

4. A method has been developed to adaptively focus an electron beam during EBW. The method provides highly accurate control of the focus of an electron beam during welding with an oscillating focus in conditions of deep weld penetration. The method is based on the use of synchronous demodulation and makes it possible to determine and adjust the focusing current of an electron beam during EBW, which during the welding process provides either maximum weld penetration of the metal (a sharply focused beam) or a penetration value close to the maximum (over-focused or under-focused beams).

References

[1] R. Rai, P. Burgardt,J.O. Milewski, T.J.Lienert, T.DebRoy, "Heat transfer and fluid flow during EBW of 21Cr–6Ni–9Mn steel and Ti–6Al–4V alloy,"*J. Phys. D: Appl. Phys.*, vol. 42, no. 2,pp. 1-12, Jan. 2009.

[2] R. Rai, T.A. Palmer, J.W. Elmer, T. DebRoy,."Heat transfer and fluid flow during EBW of 304l stainless steel alloy,"*Weld. J.,*vol. 88, pp. 54s-61s, March 2009.

[3] VidyutDey, Dilip Kumar Pratihar, G.LDatta, M.N. Jha, T.K. Bapat, "Optimization of bead geometry in EBW using a Genetic Algorithm,*". J. of Mat.Proc.Ttech.*vol. 209, iss. 3, pp. 1151–1157, Feb. 2009.

[4] E. Koleva, "Electron beam weld parameters and thermal efficiency improvement",*Vacuum*, 2005, vol. 77,iss. 11, pp. 413-421, March 2005.

[5] "Recommended practices for EBW," ANSI/AWS, American Welding Society, Miami, FL, 1999, pp.1-99,

[6] J.W.Elmer, A.T.Teruya, "An enhanced faraday cup for the rapid determination of the power density distribution in electron beams,"*Weld. J.*, vol. 80, no. 12, pp. 288-295, 2001.

[7] T.A. Palmer, J.W.Elmer, K.D.Nicklas, T.Mustaleski, "Transferring EBW parameters using the enhanced modified faraday cup,"*Weld. J.*, vol. 86, pp. 388s-398s, Dec. 2007.

[8] I.A. Krinberg, G.M. Mladenov, "Formation and expansion of the plasma column under electron beam-metal interaction," *Vacuum*, vol. 77, pp. 407-411, 2005.

[9] D.N.Trushnikov, V.Ya.Belen'kiy, V.M.Yazovskikh, L.N.Krotov, "Formation of a secondary-emission signal in EBW with continuous penetration,"*Weld. Int.*, vol. 21, no. 5, pp. 384-386, 2007.

[10] V.M.Yazovskikh, D.N.Trushnikov, V.Ya.Belenk'iy, "The mechanism of secondary emission processes in EBW with the modulation of the electron beam,"*Weld. Int.*,vol. 18, no. 9, pp. 724-729, 2004.

[11] D. N. Trushnikov, "Using the wavelet analysis of secondary current signals for investigating and controlling EBW"*Weld. Int.*, DOI:10.1080/09507116.2012.715932, Oct. 2012, to be published.

[12] V. Zuev, N.N. Rykalin,A.A. Uglov,"Fundamentals of electronbeam processing of materials,".*Moscow: Mashinostroenie*; USSR, 1978. p. 239 [in russian].

[13] W. Teng, G. Xiangdong, S. Katayama, J. Xiaoli,"Study of dynamic features of surface plasma in high-power disk laser welding," *Plasma Science and Technology*, vol. 14, no. 3, pp. 245–251, 2012.
[14] A.F.H. Kaplan, P. Norman, I. Eriksson,"Analysis of the keyhole and weld pool dynamics by imaging evaluation and photodiode monitoring,"*Proceedings of LAMP2009 – the 5th Int. Congress on Laser Adv. Mat. Proc*., pp. 1–6, 2009.
[15] S. Teresa, A. Antonio, R. Domenico, L. Valentina, T. Luigi, M. Pietro,"Plasma plume oscillations monitoring during laser welding of stainless steel by discrete wavelet transform application,"*Sensors,*vol. 10, no. 4, pp. 3549–3561, 2010.
[16] Y. Peng, W. Chen, C. Wang, G. Bao, Z. Tian,"Controlling the plasma of deep penetration laser welding to increase power efficiency,"*J. Phys. D: Appl. Phys*.,vol. 34, no. 21,pp. 3145–3149, 2001.
[17] J.Max, "Méthodes et techniques de traitement du signal et applications aux mesures physiques. 3rd ed.,";*Publisher: Jean-Louis Lacoume, Paris* 1981; vol. I, pp 10-50.
[18] D. N. Trushnikov, V. Ya.Belenkiy, G. M.Mladenov, N. S.Portnov, "Secondary-Emission signal for weld formation monitoring and control in EBW (EBW),"*Materialwissenschaft und Werkstofftechnik*, vol. 43, no. 10, pp. 892–897, Oct. 2012.
[19] S.J.Orfanidis, "Optimum Signal Processing. An Introduction. 2nd Edition,"*Publisher: Prentice-Hall*, Englewood Cliffs, New York, 1996; 5-30.
[20] I.Y. Smurov,A.A. Uglov, A.M. Lashyn, P. Matteazzi, L. Covelli and V. Tagliaferri,"Modelling of pulse-periodic energy flow action on metallic materials,"*Int. J. of Heat and Mass Transfer*, vol. 34,pp. 961–971, 1991.
[21] D. Trushnikov, V. Belenkiy, V.Shchavlev, A.Piskunov, A.Abdullin, G. Mladenov, "Plasma charge current for controlling and monitoring EBW with beam oscillation ," *Sensors*, vol. 12, pp. 17433-17445, Dec. 2012.

Methodical Bases of the Integrated Electrotechnical Complexes Life Cycle Logistic Support

Anton Petrochenkov
Perm National Research Polytechnic University - Electrotechnical department
Komsomolsky avenue 29, 614990 Perm, Russia
E-mail: pab@msa.pstu.ru

***Abstract*—Complex aspects of management by life cycle of electrotechnical complexes of the oil-extracting enterprises by a method of the integrated logistic support are considered.**

***Keywords:* life cycle; electro-technical complexes; technical condition; integrated logistics.**

I. INTRODUCTION

OIL industry is the key one for the economics of Russia Federation. Taking into account that the cost of oil production in Russia is high relative to most other oil producing countries, and the expenditures for electricity and maintenance of the energy sector are the most significant, (from 30 up to 50%), the task for developing measures as well as software and hardware tools to improve the usage of electrotechnical complexes (ETCs) of oil industry (OI) is very important.

Valid till the 90s documents and statutory documents, industrial standards fixed high level of control maintenance, repair, and operations system (MRO) management system in the oil industry at that times.

But unfortunately the situation was changed in the last years. The reached positions weren't developed and as a result were lost. The change of social and economic structure of the country brought to the reforming and reorientation of the business processes of OI (with the orientation to the "building in the world economics"), but it didn't affect the engineering and social responsible aspects. Service companies (also so called energy service companies), providing the support of the life cycle (LC) electrotechnical complexes of the oil companies, are in operation apart from the OI, the goals of their business processes are mostly different. The existing disadvantages in the part of creating the industrial relation system are acquiring in decreasing the efficiency and reliability of electrical engineering system [1].

No doubt, a number of international guidelines and standards describing the principles of modern life cycle management systems of different objects; the MRO task is considered by the leading enterprises as part of the whole life cycle management tasks, taking into account their engineering- social-economic aspects. At the same time, we must take into account the following factors:

First, foreign standards are making the feedback "object MRO- configuration MRO" on the base of analysis of fault operating data, but not foreseen the usage of the MRO results for chancing the MRO system without having the failure, but the scientific based solution of the problem is not being opened. The appearance of failure fact in the corresponded exploitative documents of the enterprise is rare situation, because of the existing preventive maintenance service and because of relations between the enterprises and the service company. Information base of the made decisions is getting narrow, and there is a need in searching the ways of MRO applied results.

Second, for LC management in ETC of OI there is necessity in the adaption to the foreign standards (differences in exploitative documents, different reading in local and foreign composite functions of LC ETC management, supervisory control, information management complex, etc.). Also there is a problem of different

Also, there is the problem of the different requirements for schematic images and symbols of electric power system elements. Differ in terms of the standards and requirements for the provision of information and energy life cycle ETC management process.

Thus, the solution to the problem of process management of ETC OI lifecycle support in accordance with international quality standards and embedding it in a single, integrated enterprise management system requires the development of evidence-based connection between analyzed results of operating ETC OI and changes in LC system support configuration [1].

II. Methodological Foundations Of Integrated Logistics Support Life Cycle Of Electrotechnical Complexes For Oil Industry

The key task in providing the ETC LC is the support for integration of all information flows about the LC not only in the process of planning and development but in distribution, integration, and operating.

The system approach in LC planning of any object and after a complex of administrative measures, aimed at cutting the costs, are combined under the term of integrated logistic support.

ILS – methodology of object LC cost optimization in accordance with criteria of its best availability to the operation support, reliability and repair capability based on creating the integrated logistic support.

Applied to ETC OI in ILS should be considered [2]:

- Research perspectives of the ETC;
- Determination of infrastructure services of ETC during the operation, including the planning procedures logistics, diagnostics, repairs, etc.;
- Respect for the maintainability of ETC design, development of sophisticated technology services along with the development of the ETC;
-Calculation of reliability and longer uptime ETC;
- Calculation of the production costs and operation costs of the ETC;
- Determination of the composition and the quantity of spare parts;
- Staff Training;
- Support links between producers and consumers by consumers access the integrated database of equipment in order to simplify its diagnostics and repair, as well as obtaining the manufacturer of faults and failures in order to take measures to improve the reliability of the ETC;
-Classification and codification of goods and materials needed in simplification the search of the right data in directories and databases (DB) to avoid duplication of projects, accelerating the preparation of applications for supply of components, etc.;
-Development and maintenance of exploitative documentation;
-Traditional logistic procedures (packing, storage, transportation, etc.).

The system of integrated logistic support (ILS) of the ETC of the OI is the integrated logistic system, providing support of the ETC operation during the whole its LC in accordance with the ILS specifications and industry characteristic of the enterprise.

Correspondence of the ETC LC stages and LC of the ILS is shown on the Fig.1.

Engineering stage of the ETC is equal to the ILS system engineering stage (engineering the strategy of operating support). The method of successive approximations defines structural features of ETC, providing the best availability for the support service. Each option has its own design and functional characteristics, respectively, the characteristics of ILS.

On the operating stage (support of ETC operation) the monitoring, diagnostics, analyses of ETC conditions and its ILS system are being carried out in order to check the adequacy of the measured and predicted value of its characteristics. Discovered imbalance between measured and predicted value can lead to the revision of the ILS system, and in the worst case scenario it can lead to the revision of the functional features and even the whole concept of the ETC.

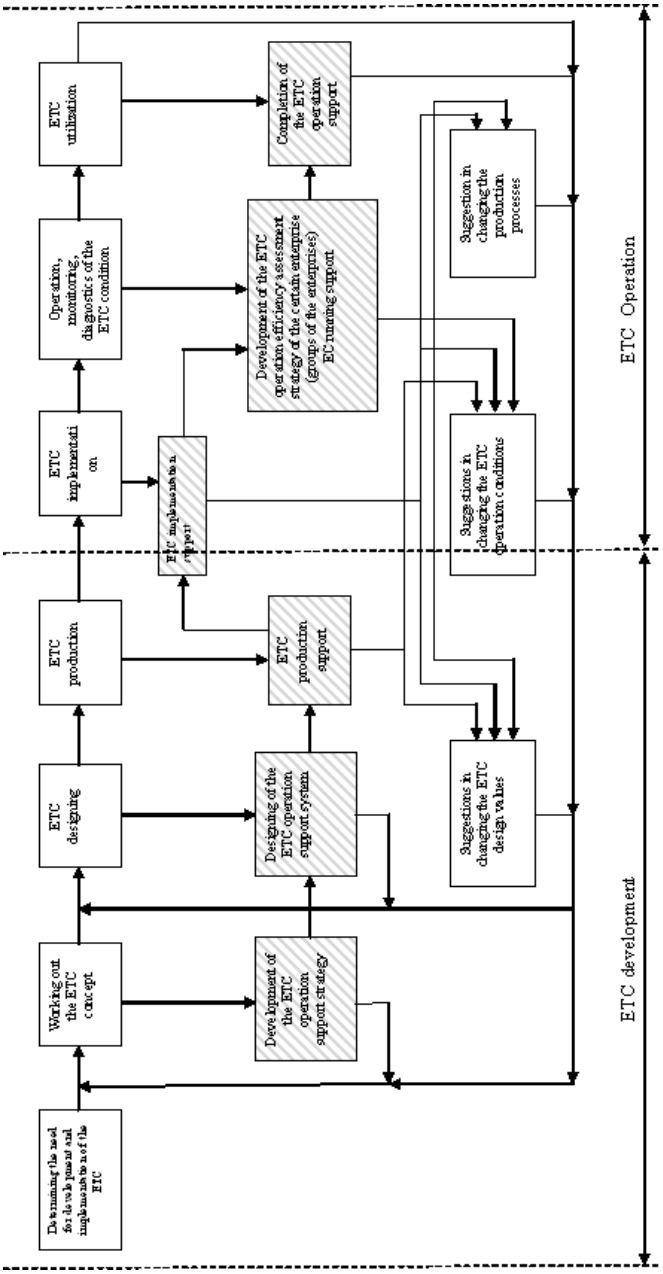


Fig. 1. ILS and ETC Life cycles.

After utilization of the ETC the total cost is being calculated of the ETC LC and the total efficiency of the ILS organization is estimated. Such estimation together with historical data about the ILS system can be used during the organization ILS for the ETC of the same type or use.

The main thing one should pay attention on is the

permanent control for the current cost and real cost of the ETC LC. The importance of the strict control is the necessity to prove the enterprise-exploiter that the got cost of the ETC LC is optimum [2].

ILS includes the following procedures:

1) ETC Logistic Support Analysis, which is run with the aim to provide with necessary reliability level, repair capability, available to support and requirements determination:

- to individual components of the ETC items, placing units and units subject to regular maintenance, replacement and repair;
- to the support and test equipment;
- to the number and qualifications of operating and maintenance personnel;
- to the system and training facilities;
- to the range and quantity of spare parts, consumables, etc.
- to the organization of storage, transport, packaging, etc.

2) Maintenance, Repair, And Operation Planning:

- Development of the MRO concept, requirements to the ETC in its maintenance and implementation of the MRO plan;

3) Integrated Supply Support Procedures Planning – integrated procedures of logistic support of the operation processes, maintenance and repair including:

- Determination of the parameters of the initial and current logistics;
- Codification of supply items;
- Planning of products supply;
- Management of orders for supplies delivery;
- Management of payments ordered provisioning items;

4) Measures in providing the staff with electronic maintenance documentation, electronic repair documentation, which are held on the stage of designing and implemented in the process of certain assembly line of the equipment. Mentioned documentation is used during the purchase, transferring, supplying, implementation, during the operation, maintenance and repair of the electrotechnical complex.

One of the most important phases in terms of a given program ETC life cycle, as well as taking into account the specific characteristics of the enterprise, is the stage of the equipment operation.

The main types of impacts and support of the ETC OI operations are the following [3]:

1) Changes in the ETC operating mode (change operating procedures and power reduction);

2) ETC upgrading (with a second vehicle diagnostic systems, improve operational parameters);

3) Change in the ETC strategy maintenance.

Evaluation of the ETC OI technical conditions is an important task, which can detect the early stages of incipient defects and, thus, to prevent accidents that can result in serious adverse effects [2]. Instead of the current system in most enterprises of preventive work, which includes periodic monitoring and prevention, more appropriate, from a technical and economic point of view, is to serve the ETC on its actual state. It means that repairs are made only when the need arises.

Therefore, a comprehensive strategy must be built in the frame of the system to ensure a given technical condition of the ETC [4].

In general the task of ILS LC ETC OI includes several stages:

Definition of the elements of power supply system (PS) based on technological inter-linkages;

Identification of the external (system) constraints;

Determination the needs of specific elements of the ETC in the amounts and timing of repair;

Definition of internal restrictions of the timing and costs (including all software repair);

Coordination of internal and external constraints;

Repair schedule, which includes timing and duration of outages.

The essence of this problem is to determine the ETC repair programs using both regulatory and technical condition assessment results (TC). This combination of information allows you to make full use of any information on the equipment, received by the diagnosis systems and due to expert's assessment [5].

Repair program means schedule of the specific sets of repairs, fully reflecting the dynamics of change in technical condition equipment.

Diagnosis and monitoring of the technical condition (TC) of the equipment involves measurement of different physical quantities: partial discharge, electrical leakage, return voltage, acoustic waves, vibration, infrared light, etc.

One needs to highlight the problem in information management. A huge amount of information influences the efficiency information management about ETC OI (there is so called "information chaos"). That is why it is necessary to create the common information space (CIS). The considered group CALS-technologies consists of known methods, which can be divided into 3 groups [1]:

1. EC designing (on the base of CAD/CAM/CAE-technologies);

2. EC production (on the base of technologies of automation processes and its planning);

3. Delivery, implementation and operation of the ETC (technology-based automation processes of the supply and use of the product), in particular:

3.1 ETC logistic support system;

3.2 E-trading systems;

3.3 Interactive electronic technical manuals (IETM) (automated systems that provide operational information to the user on a specific equipment and the ability of the hardware diagnostics, troubleshooting, training, etc.).

Analysis of the current situation in the industry enterprises showed that there is a lack of both statistical and operational, regulatory, guidance documentation about ETC and its elements, and there is a low degree of efficiency of its search.

In the frame of this approach there were developed IETM for the enterprises OOO LUKOIL–Permnefteorgsintez and OOO LUKOIL-Perm on the base of Microprocessor Units Automatization Chair, State National Research Politechnical University of Perm.

Methodological complex, providing the ETC OI LC

management by the ILS (example, 643 units of the package transformer substations 6/0,4 kV of the oil and gas workshop) was developed.

For providing the technological conditions (TC) the survey in the first year two are held, the next year - one by one (the cycles of control TC 1 time per 4 years) - only n = 5 times. The calculations on the number of one shift during the research of PTS are also done.

The analysis for implementation of managerial mechanism of the ETC OI LC was done and suggested the managerial structure of power divisions of oil industry. The scheme illustrated the TO logistic system and also was illustrated the scheme of ETC OI LC elements (objects) decision making is shown on the Fig. 2.

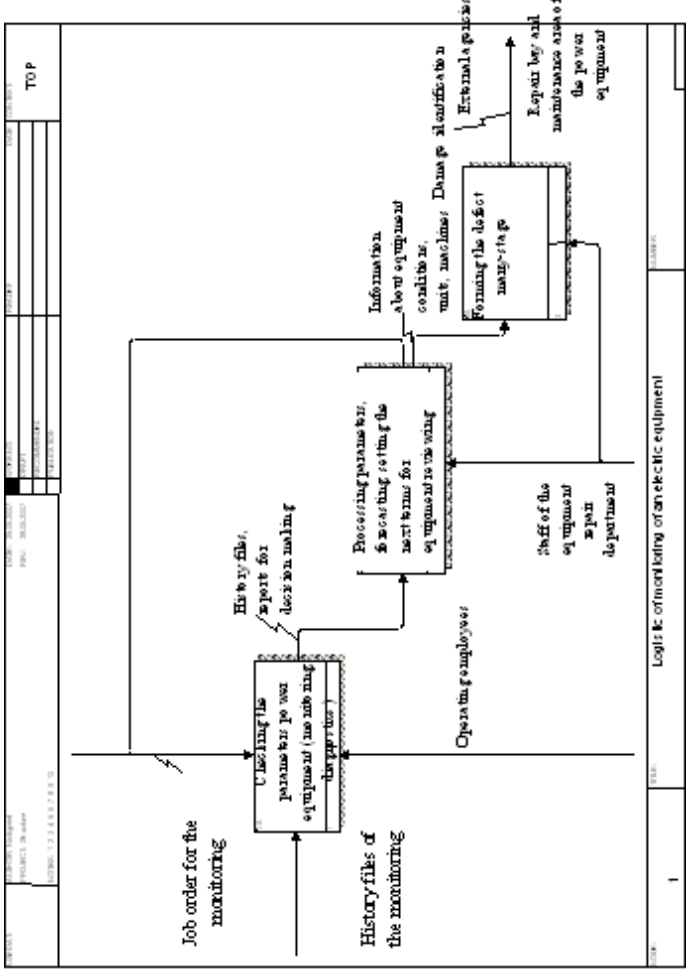


Fig. 2. The scheme illustrated the TO logistic system and also was illustrated the scheme of the elements of ETC OI LC decision making.

Key factors for the company in this matter are time monitoring and payments for its holding. The economic effect of the implemented system was significant and accounted for about 16% over the period of functioning in 4 years [6].

Methodological provision of ETC equipment control elements is developed. For example, a model of staff behavior at operation of gas protection is shown on Fig. 3.

It is necessary to note that the given procedure of subjective judgment and decision-making has positive experience of application by the commission of experts in the field of electric power industry: it was successfully applied when choosing the main schemes of electric connections, as well as schemes of own needs and electrical supply schemes.

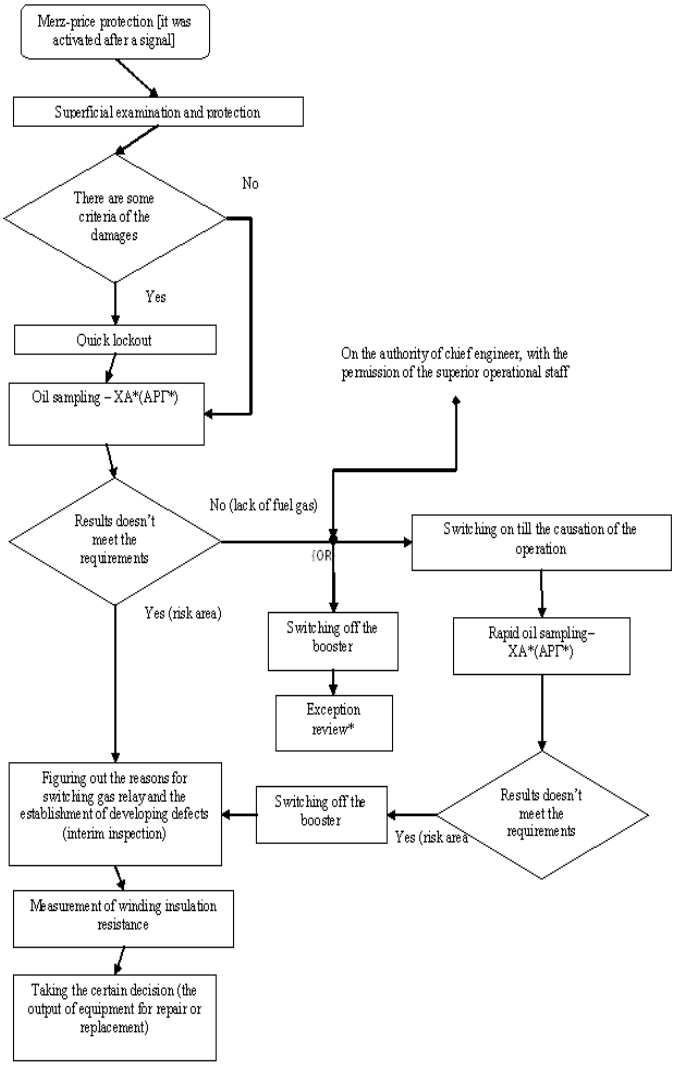


Fig. 3. Flow scheme of the maintenance staff response actions upon activation of gas protection (a fragment of the developed method).

The particular problem of money resources distribution of complex works according to the condition high-voltage electro technical equipment system (some key parameters) can be shown to expert ranging of variants of electrical supply systems [1].

In general, contemplating upon the problem of the information of technical conditions of the ETCs use for the decision-making, it is rational to allocate the following levels of adequacy of estimations:

- Identification of the technical condition with the help of reliability indicators, i.e. parameter of the refusals or intensity

restarts stream.

- Identification of the technical condition with the help of likelihood characteristics of defects and damages revealed during the certain moments of time.

- Identification of the condition with the analysis of the continuously controllable technological parameters that characterize the technical condition of the equipment elements.

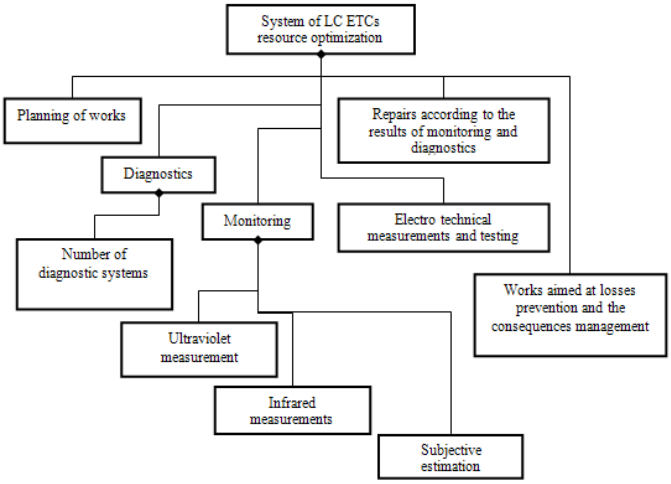


Fig. 4. Optimization principle of the resource of LC ETCs equipment system.

Realization of the third level is possible only under the condition of perfect diagnostic systems and adequate mathematical models of correlation communications between the target parameters and the technical condition of the equipment. The second level is more accessible and demands the sufficient information on the defects and damages revealed, as well as dynamics models of their development. Realization of the first level is connected only with the sufficiency of the statistical material.

The correctness of the decision-making on the introduction of the particular regulations of the maintenance service and repair depends on the technical condition of the equipment. The maintenance service and repair regulations as the system of the rules defining technology, means, volume, methods and periodicity of repair influences, depend not only on a technical condition of the considered equipment. It is influenced by the structural importance of the given equipment in the comprehensive technical system as well as concrete conditions of its functioning (including purely repair character, for example, equipment of repair base, presence of resources, etc.).

Three levels of identification of a technical condition mentioned above, as a matter of fact, are the three levels of an estimation of reliability function. The first level corresponds to the zero approach when the aprioristic information is absent. The second level corresponds to the case when aprioristic casual process of accumulation of damages of the given type is set. The third level concerns the case when the forecast is carried out separately for each realization of the casual development of refusal. Use of this or that level is defined by the necessity of the forecast accuracy and presence of corresponding means, program and supply of information [2].

The optimization principle of the resource of LC ETCs equipment system is shown on Fig. 4.

III. Conclusion

The proposed scheme corresponding to the stage of the life cycle of ETCs and stages of integrated logistics support will develop rules for the ETC documentation design based on selected ETC configuration.

There was developed methodological support for the life cycle management of key engineering equipment of ETC extraction equipment on the basis of integrated logistics support for the enterprises OOO LUKOIL–Permnefteorgsintez and OOO LUKOIL-Perm.

Electronic educational resources were developed using an educational process for training students with the specializations “Electrical Power Supply,” “Automation of Technological Processes and Production,” and “Automated Management of Product Life Cycle” of State National Research Politechnical University of Perm [7].

References

[1] A. B. Petrochenkov, “Regarding Life-Cycle Management of Electrotechnical Complexes in Oil Production”, *Russian Electrical Engineering*, vol. 83, no.11., pp. 621-627, 2012.

[2] S. V. Bochkarev, A. B. Petrochenkov, A. V. Romodin, *The integrated logistic support of operation of electrotechnical products*. Perm, Russia: Publishing house of Perm State Technical University, 2009, pp. 283-294.

[3] A. B. Petrochenkov, A. V. Romodin (2011). Energy-optimizer complex. Russian Electrical Engineering. [Online]. vol. 81, no. 6, pp. 323-327. Available: http://link.springer.com/content/pdf/10.3103%2FS106837121006009X

[4] A. B. Petrochenkov, A. V. Romodin, N. I. Khoroshev, “About one formalized method of an assessment of administrative decisions (on an example of management of electrotechnical objects)”, *Scientific and technical sheets Saint Petersburg State Polytechnical University*, no. 5 (87), pp. 166-171, 2009.

[5] *Volume and norms of tests of electric equipment*, RD 34.45-51.300-97, 2001.

[6] *System of maintenance, control of a technical condition and repair on a technical condition of the oil-field equipment of affiliated oil and gas extraction societies of JSC Lukoil*, STP-01-026-2003. JSC Lukoil standard, 2003.

[7] A. B. Petrochenkov, A. V. Romodin, B. V. Kavalerov, L. A. Mylnikov, V. P. Kuprin, *Database and control system of the Energy Data Ware database*. Perm, Russia: Publishing house of Perm State Technical University, 2008, pp. 260-326.

Issues of information support for the process of scientific enterprise's management in the universities and research organizations

Leonid Mylnikov, Alexander Kalikh
Perm National Research Polytechnic University - Electrotechnical department
Komsomolsky avenue 29, 614990 Perm, Russia
E-mail: leonid@pstu.ru, kalikh@bk.ru

Abstract - **This article describes the problem of commercializing of scientific researches in universities. Management tasks are reduced to subtasks and combined formal algorithm. The overall control problem is reduced to a set of formal subtasks combined into a single algorithm. Here the necessity of joint control of all commercialization projects as well as the use of information systems for the successful implementation of the existing commercial potential is shown.**

Keywords - innovations, management, algorithm, support, information

I. INTRODUCTION

Nowadays, a lot of attention is being paid to issues related to the so-called high-tech, scientific enterprises. Enterprises where new scientific researchers are embodied in new products and can significantly improve existing products or technologies. Sometimes, such developments may create new markets. About the most successful examples a lot of books were written (personal computer, cellular phone, television, car).

Reasons for the rise and fall of individual civilizations and the impact on these processes of new technologies are being researched by the historians (the wheel, bronze, iron, etc.).

No one have any doubt about the fact that the impact of technology on our lives is enormous, and that its sources are research teams and scientists. However, despite the fact that such an important indicator of academic performance as the number of publications worldwide annually grows, the similar increase in the number of innovations embodied in the specific product or products does not occur. Particularly low number of such realizations is carried out in universities where new research results appear.

At present, universities and research organizations have a limited number of ways to implement its developments: for sale of licenses, for implementation of applied researches on the orders of large corporations, the creation of enterprises by the university or its employees. The gap between the amount of funds received by universities for researches and earned by universities is huge. For example, it was spent $ 48,164,473,678 for the university science in the U.S. in 2009, and only $ 1,782,113,228 were received for licensing. That is only 3.7% of the money spent (The Chronicle of Higher Education 2010). And it's, despite the fact, that the U.S. is considered the most successful country in terms of earning money for university scientific research.

Customized Research, of course, brings a significant contribution to the budget of many universities. However, holders of such developments are the customers, and customers are concerned with the further promotion of the results (Prager and Omenn 1980, pp 379-384). Therefore, research organization, in this case, prevented from receiving the profit from such research and development.

Entrepreneurship today is carried out in such a way that management of the University not only doesn't prevent but also assists professors who want to create a firm(at Harvard University, for example, professors are granted for paid vacation for a period of six months). Companies that wish to purchase a license for a particular patent or know-how as a rule do not refuse.

Managing the process of commercialization is currently focused mainly on patenting of scientific developments, commercial evaluation and finding ways to commercialize it (the sale of the license or the establishment of the firm in this case the further fate developments is in the hands of entities that received them). It is believed that large companies are experienced participants in the market and would not allow for technologies not remain in demand, and new enterprises should gain the experience of entrepreneurship (by creating and bankrupting itself until they find their niche). Such a strategy of small firms is not acceptable in some countries such as Russia. Because of the legislation of many developing countries to create a new firm is often much simpler and cheaper than to close it. In addition, each research group has its own specialization and thereby creates technologies that are often competing and the university itself creates its own ob-

stacles, to make a profit producing such technology to the market in different ways.

The obvious solution to the problem of improving the process's management of earning money through the commercialization of universities is the integrated management of the process.

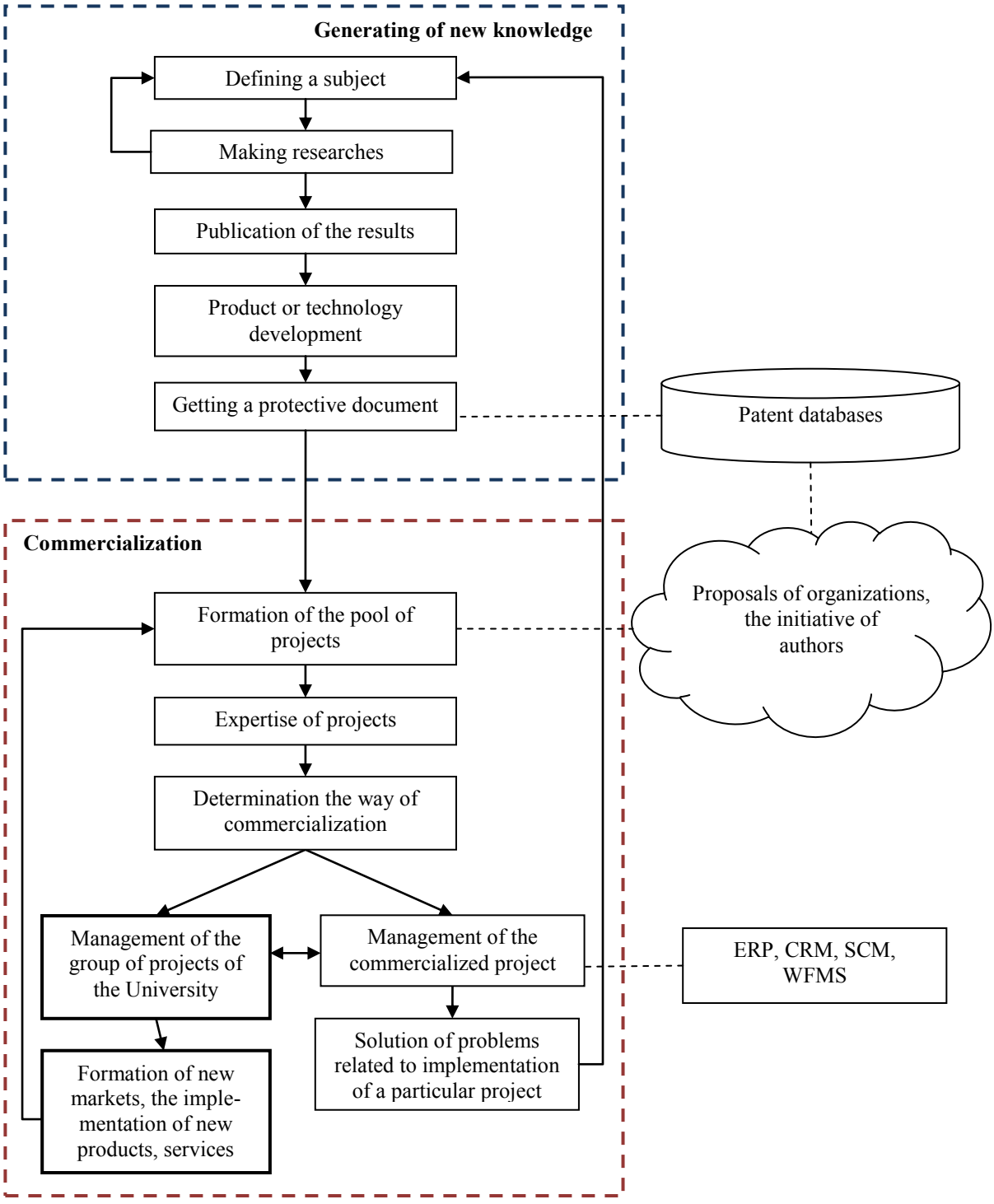


Figure 1. The system of commercialization of scientific developments from the perspective of the university as a founder of new innovative enterprises

II. STATEMENT OF PROBLEM

Management of any process based primarily on an understanding of the principles of their occurrence.

Understanding of what is happening tightly associated with the degree of formalization of process's components.

It should be understood that the management of an entire group of companies will be held as management of a single corporation (as the management of individual enterprises does not bring the desired effect). In connection with this management process will include steps which are not used in managing individual enterprises, and which currently do not have the instrumental support for their implementation (Fig. 1). Moreover most of steps in this process are so complex that their formalization is too difficult (except on qualitative level).

In addition, product-oriented processes (Stadler 2011, pp. 44-62) associated with the commercialization scientific projects, usually occur within a limited period of time and have strongly pronounced stages of development, which are limited in time (Schwarzer and Krcmar 2010). Each project comes to the stage of stabilization. For the majority of innovative projects (projects of commercialization of scientific research) the transition to this stage does not make substantial profits or may even be not profitable. In addition stable income-generating project no longer requires special attention because it is not confronted with problems that arise during the initial stages of commercialization. Therefore, commercialization of researches is described within the project approach. This greatly increases the responsibility for the quality of decisions.

Information systems of management are widely used now in order to increase the objectivity of decision-making during the management process. For their application it is necessary to: create directories of control parameters, the classification of processes and phenomena which occur in the management, standardization of management processes, the development of decision-making algorithms , the study methodological aspects of management, definition of decision points.

III. THE SUPPORT OF MANAGEMENT PROCESSES

Information support in solving problems of management of innovative projects is primarily focused at supporting of decision-making.

If the generation of knowledge refers in part to the creativity, a part associated with the commercialization may well be formalized. In our context, the approach to the formalization can be considered from two perspectives: the position of existing algorithms and information systems and the positions of structuring and convenient representation of the subject area to the person who will make a decision.

A part relating to production and registration can use existing algorithms and information systems.

However, parts related to the examination and selection methods of commercialization require more detailed consideration. This is primarily due to the fact that there is no universal algorithm that can be applied. Typically, such decisions are based on the experience of the person that makes decisions and on emerging trends. Therefore, above all, structured information and methodological ways to inform decision-making are required for their acceptance.

For accepting of such decisions in case if statistics had been accumulated, graphs with the vertical axis (Müller 2011, pp. 26-29) are often used. Thus, the principle of accumulation of information and identifying key for the success of the project values or ranges of values can be implemented (Figure 2, Tab. 1.).

Thus, as shown the structuring of information, even in small quantities allows you to emphasize, if not the corridors of values, which often lead to a result, then values which often lead to failure. For example, Fig. 2., shows that the chances of success falls down with the assistance of a second university or research organization, with the use of certificates of registration of computer programs and databases with estimates of the cost of > 30 thousand rubles. (Especially in IT).

Determination of the way of commercialization does not end when it is decided to sell the license or to establish the company, and the form of ownership is defined. The important thing is the positioning of a potential product in the market. It is here to show the "corporate" in the management process. It should take into account the effects of other projects, which realizes the university, and to find a niche in this situation. So-called portfolio matrix may be used here (Tab 2). It defines four possible structural concepts for the development of innovative trends or a separate product.

The first concept, which is reflected in the matrix, describes the behavior of the company relatively to the developed products, implemented in existing markets. The innovative enterprises' behavior in this situation should be directed to the intensification of use of existing market’s possibilities by positioning of already known products on them. That means to ensure the competitiveness of products in the markets. If the product is in its characteristics do not appear perspective then the innovative position of the enterprise may be manifested in means of removing it from the market in order to open a niche for a new innovative product.

The second concept describes the behavior of the enterprise concerning promotion of the developed products to new markets. This is due to innovative activity aimed at finding of new ways of application or on searching for ways of using an existing product. This concept requires a research to create a new consumer product with the properties that take into account the requirements of new markets and the corresponding variations in the production program. Innovations in this case realized in the form of researches and developments, expanding the scope of influence of the product, invention or technology.

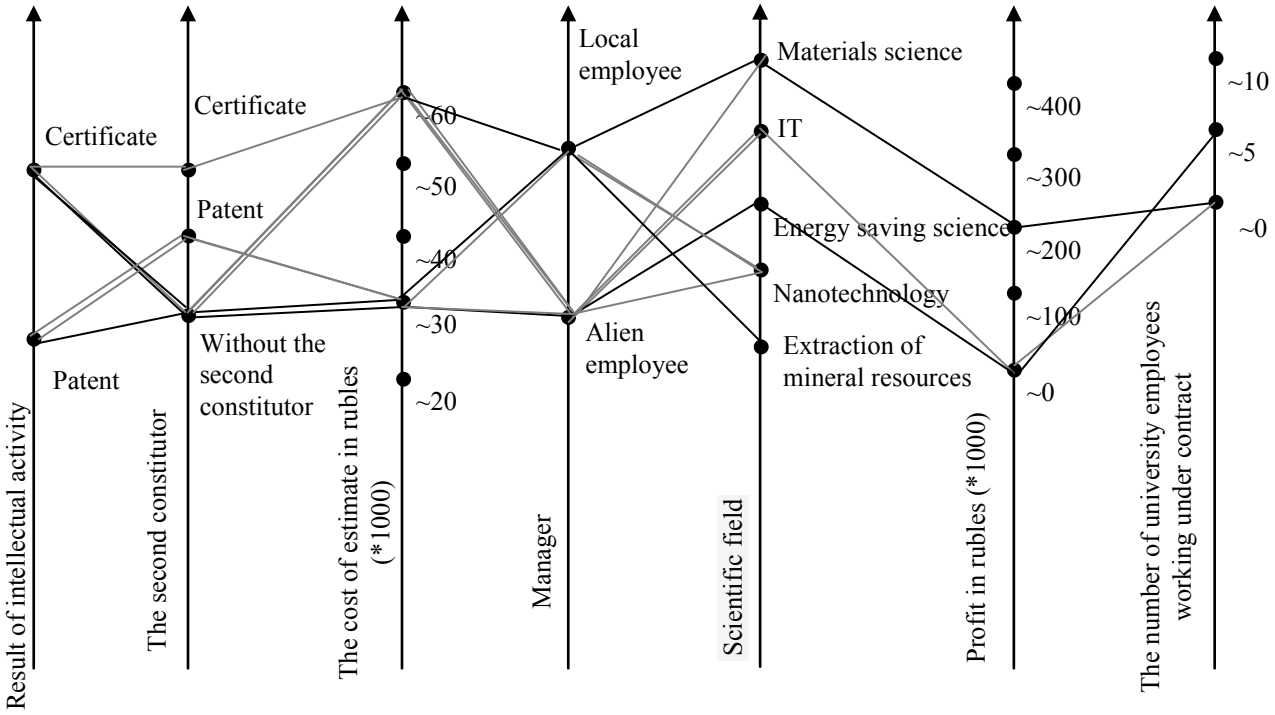


Figure 2. Graph with the vertical axis with indicators of companies created with the participation of Perm National Research University for 2010-2011 (companies, which activities are considered to be successful are marked with dark color)

The development of research and development to create new competitive product modification is the main direction of innovation enterprises in this situation. Here we speak about a work aimed at creating and developing new products or the modifications of the old products, which would have new consumer properties. Activities implement in the form of applied research and development, expanding the scope of the product, inventions or technologies.

The fourth concept (and the most productive) is associated with the creation of a new product and promoting it to the new markets. This innovative policy of the company is called diversification, i.e., expansion of research or production profile of the enterprise through the creation and development of production quality of new products and new markets for its implementation. This is the most difficult, risky and most promising concept of innovation development of capacity-building success.

This formulation of the problem determines gaps in the commercialization strategy, identifies the missing elements for a successful project, and moves from consideration of individual projects to managing a group of projects within the framework of existing models of managing. For example, group of algorithms called Slope One (Lemire and Maclachlan 2005) - for finding joint product niches, product planning algorithms (Mylnikov 2011) - for control of production resources and formation of a portfolio of projects; prediction algorithms (Mylnikov and Alkdirou 2009, pp. 293-307; Amberg and Mylnikov 2009, pp. 11-15) - to identify the most promising projects. Thus, the problem of co-management of firms comes to the problem having the mathematical formalization, and it means that here may be an algorithm which implements the principles of university management by knowledge-intensive firms (Fig. 4).

Table 1. Statistical information about the firms founded with participation of National Research Perm Polytechnic University (Russia) in 2010-2011.

Name	Result of intellectual activity	The second founder the university or institute of RAS	The cost of estimate (rub.)	Manager (staff member of the University or not)	Scientific field	Number of employees within a year after the	Amount of profit one year after establishment (thou-	Number of university staff engaged under contracts
Economic society 1	Patent	-	25 150,98	Local	nanotechnology	1	250	-
Economic society 2	Certificate	-	55 800	Alien	Information technologies	1	-	-
Economic society 3	Certificate	-	25 000	Alien	Energy saving science	4	-	4
Economic society 4	Patent	-	56 220	Alien	Information technologies	-	-	-
Economic society 5	Patent	-	25 000	Alien	Nanotechnology	-	-	-
Economic society 6	Certificate	Certificate	56 220	Alien	Nanotechnology	-	-	-
Economic society 7	Patent	Patent	25 000	Alien	Nanotechnology	-	-	-
Economic society 8	Certificate	-	54 300	Local	Extraction of mineral resources	-	-	-
Economic society 9	Patent	Patent	25 000	Local	Nanotechnology	-	-	-

The disadvantage of this approach is the fact that in the management of existing firms (such as implementing a new project through them),it must be considered the current situation within them. Each company will have its own strengths and weaknesses, which can be identified by the use of PERT or a SWOT analysis. Thus, when the strategy of implement innovative projects is developing, the internal characteristics of organizations will be considered (Utterback 1974, pp. 620-626).

The University can be considered as a management company (like the venture capital), and the formalization of the process management allows you to do tasks of developing information systems of administrative activity automation. Thus to approach to the problems of standardization of management in the given subject domain. As a result, we can expect that the percentage of applied research projects which have reached the implementation phase, as an innovative projects, will increase as well (currently only 10-15% of such projects reach the stage of commercialization).

Table 2. Product-market matrix of an innovative portfolio (Mylnikov 2009)

Products/ Markets	Existing markets	New markets
Existing products	The intensification of the markets: - Positioning of the product - Reservation of a market niche	The development of markets: - Variations of products - applied researches
New products	Product development: - Research and development - Modification of of products	Diversification

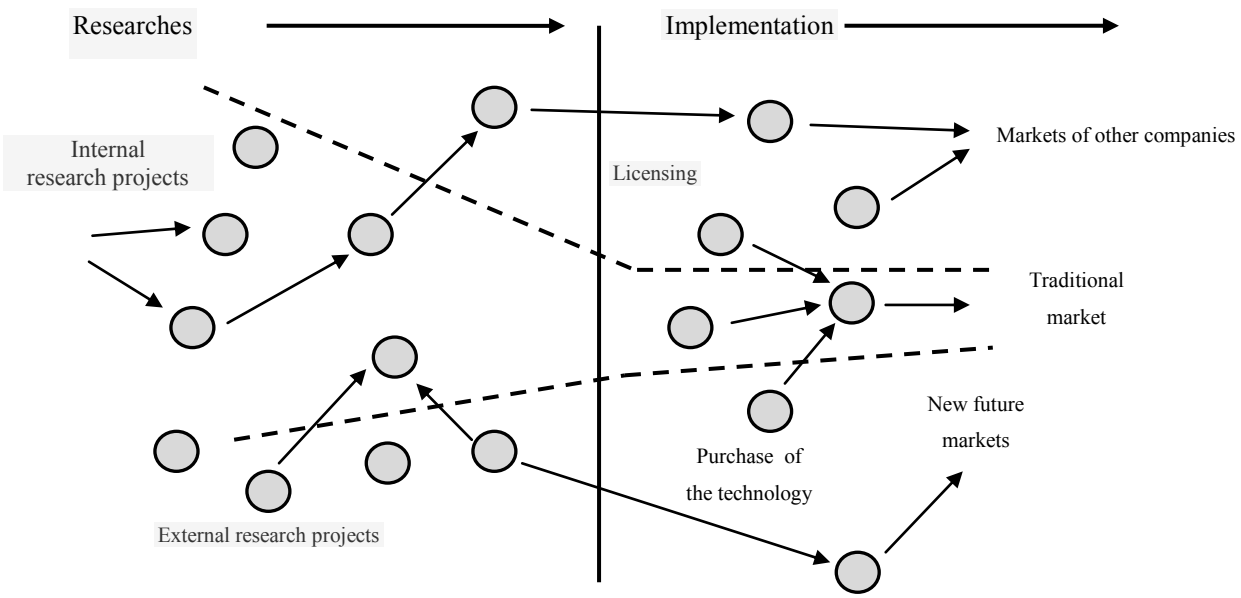


Figure 3. The paradigm of open innovation (Schwarzer and Krcmar 2010)

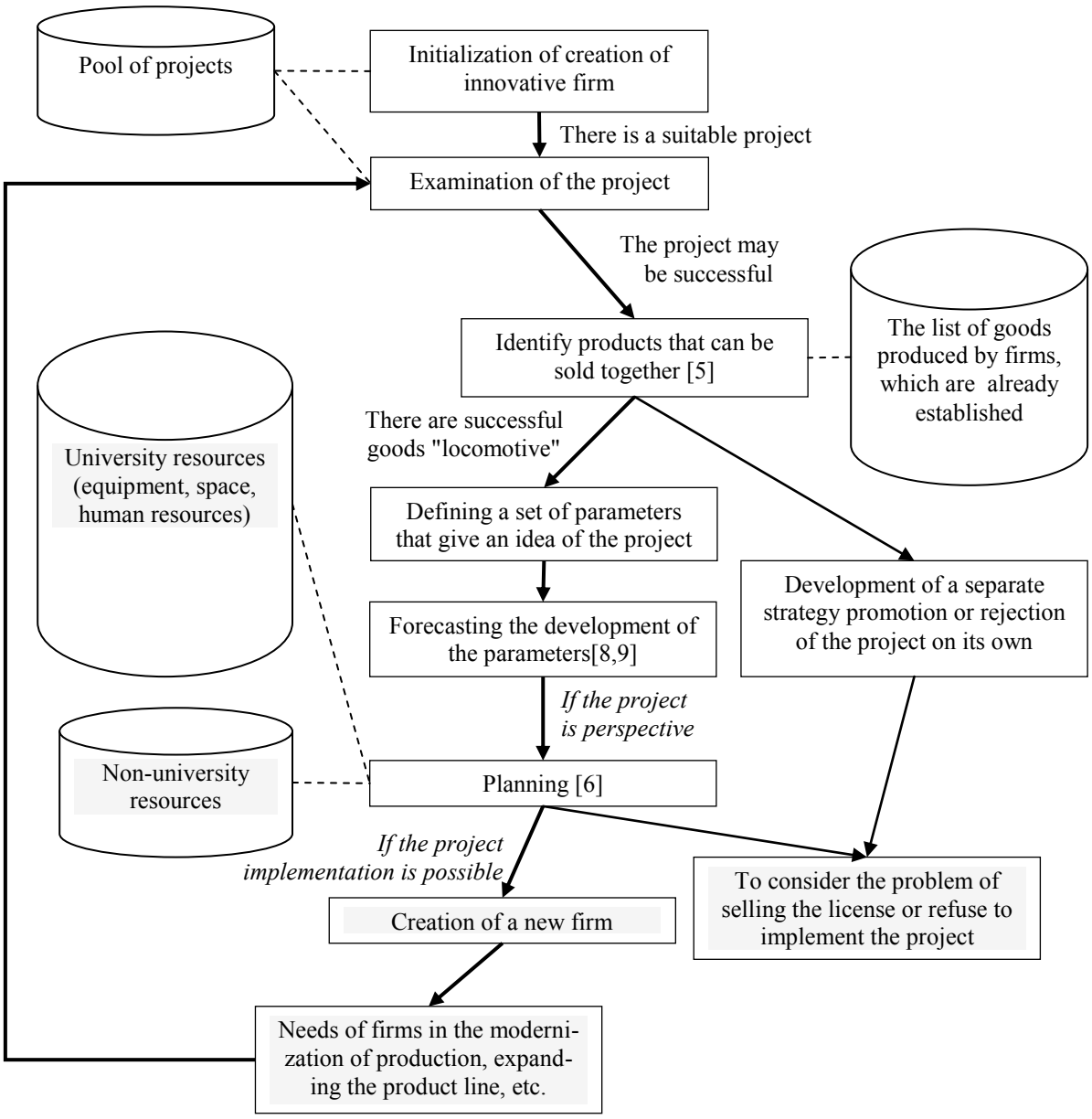


Figure 4. Algorithm for the coordination of existing innovative enterprises and creation of new, with the participation of universities or research institutions

IV. Management Information System

As follows, the information system must combine elements of analytical processing and data mining, methods for solving mathematical programming problems, must allow to make the data manipulation algorithms, allow to construct and simulate business processes, to be integrated with the DBMS.. If you look at information systems which are currently on the market (Figure 5), it becomes clear that there are currently no systems that combine all the elements. The use of large, scalable systems (which also does not satisfy all the requirements) is connected with great expenditures on buying, modification, staff. If we consider further and go directly to the level of production management problems then when going to the introduction of innovative projects, problems of Industrial Engineering arise (e.g., task management and production planning) (Mylnikov and Trusov 2011, pp. 208-213).

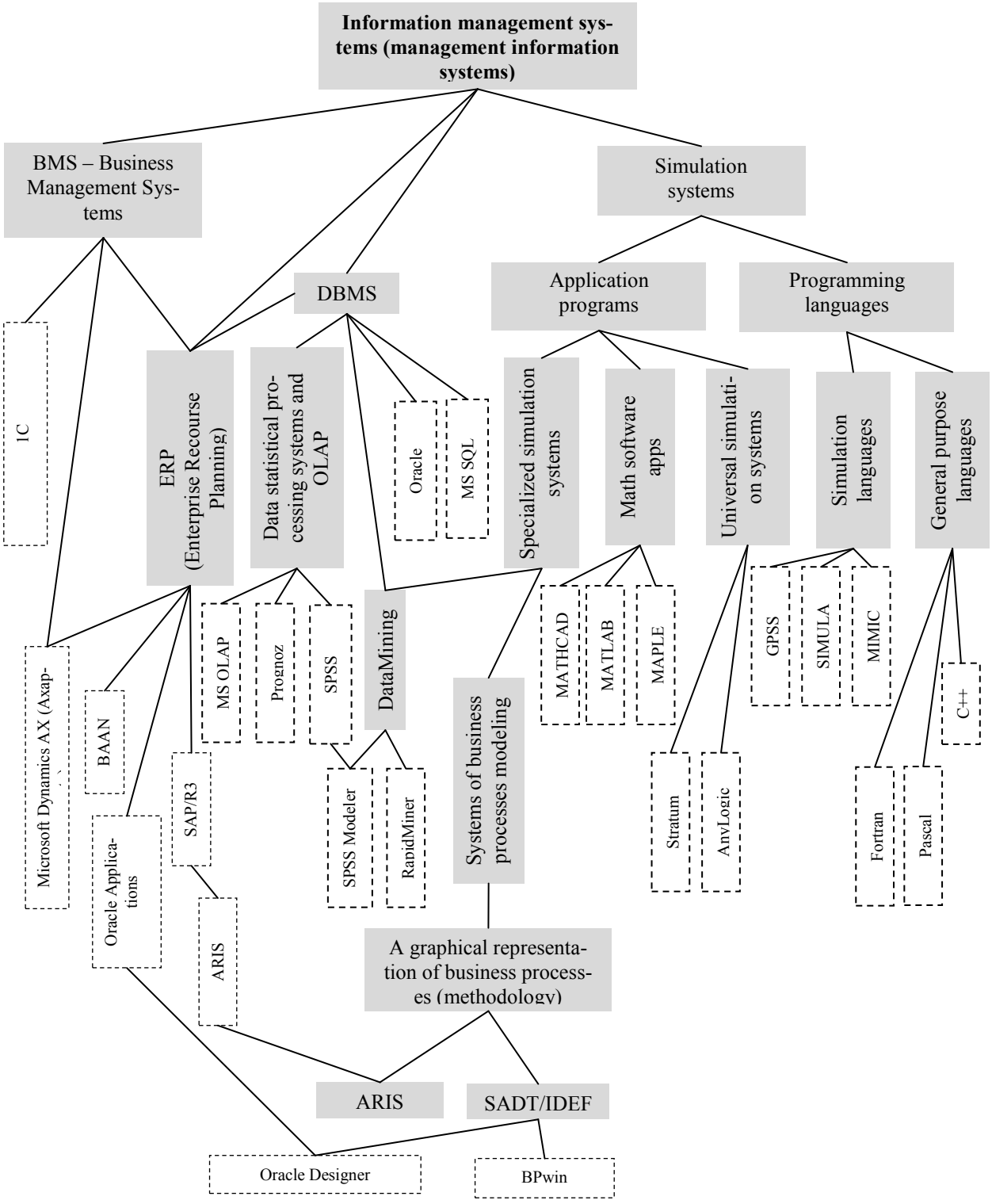


Figure 5. Approximate classification of information packets for classes of systems, the use of which is necessary for the management of enterprises according to the algorithm on Fig. 4 (gray color indicates the classes of systems, application packages and programs are marked with dots)

Adequate management can be done when it is possible to combine the two data sources: the model and data, taken from the controlled object (Shi 1987, pp. 22-29). For an adequate management of the innovative project it is required to combine together various information resources. It requires a methodical elaboration of joint use of a number of specialized IT solutions. Thus, the algorithm of coordination of implemented innovative projects can be the basis for creating a unified information system integrating software tools to support the functional stages of the innovation project. However as to create a new information system, or to integrate into a unified information space a set of information systems it should be a single information space, which can be implemented using a single bank of data (to work with this bank of data directly or through intermediate data, all subsystems should be configured . Creation of a single data bank requires the formalization of the stored information in it (in a suitable form for analysis).

V. CONCLUSION

The use of formal methods and information systems in management will help to minimize errors in decision-making, as well as develop the necessary administrative decisions, based on experience and domain knowledge. However, not all processes currently studied sufficiently. Not all processes can be adequately formalized. Key decisions will still be done by the man. Moreover, making an administrative decision, if it is based on analysis of all facts necessary for this task, does not become less complicated than the research itself results of which are planned to be commercialize. Therefore, this research itself requires additional research, the results of which can also be implemented in the form of an innovative project - management information system of innovations.

This article was prepared on the basis of of research that were performed at the Department of Applied Computer Science III University of Erlangen-Nuernberg (Nuremberg, Germany) and financed by the fund of. Hans Seidel (Munich, Germany).

REFERENCES

[1] B. Schwarzer, H. Krcmar Wirtschaftsinformatik: Grundlagen betrieblicher Informationssysteme. Schäffer-Poeschel Verlag, Stuttgart, Germany, 2010.

[2] C. Stadler Process Innovation and Integration in Process-Oriented Settings: The Case of the Oil Industry/Journal of Product Innovation Management, no. 1 (28), pp. 44-62, Jan. 2011.

[3] D.J. Prager , G.S. Omenn Research, innovation, and university-industry linkages/Science, no. 207, pp. 379-384, 1980.

[4] D. Lemire, A. Maclachlan Slope One Predictors for Online Rating-Based Collaborative Filtering. In: Proc. of SIAM Data Mining. (SDM'05), Newport Beach, California, 2005.

[5] J.M. Utterback Innovation in Industry and the Diffusion of Technology/Science, no. 183, pp. 620-626, 1974.

[6] H.L. Grob, F. Besberg, A. Coners Regelbasierte Steuerung von Geschäftsprozessen Konzeption eines Ansatzes auf Basis von Process Mining/Wirtschaftsinformati, no. 4 (50), pp. 268-281, Apr. 2008.

[7] Licensing Revenue and Patent Activity, 2009 Fiscal Year/The Chronicle of Higher Education, no. 17 (57), 2010. Available: http://chronicle.com/article/Table-Licensing-Revenue-and/125729/.

[8] L.A. Mylnikov Management of innovative projects. Perm State. Tech. Univ. Press, Perm, Russia, 2009.

[9] L.A. Mylnikov Decision support in the management of innovation projects. Perm state Tech. Univ. Press, Perm, Russia, 2011.

[10] L.A. Mylnikov, R.H. Alkdirou Approach to forecasting the development and lifecycle management of investment projects/Management of large system, no. 27, pp. 293-307, 2009.

[11] L.A. Mylnikov, A.V. Trusov (2011) On an Approach to the Design of a Logical Model of Innovation Project Data/Scientific and Technical Information Processing. No. 3 (38), pp. 208-213, 2011.

[12] M . Amberg, L.A. Mylnikov Innovation project lifecycle prolongation method/Communications of the IBIMA, no3, pp. 11-15, Jan. 2009.

[13] W. Müller Pfade im Informationsdschungel/Spektrum der Wissenschaft, no. 11, pp. 26-29, Nov. 2011.

[14] Z. Shi Knowledge-Based Decision Support System/Journal of Computer Science and Technology. no. 2 , pp. 22-29, 1987.

Best-Practice - Mobile Testing
Viewing mobile usability studies with the example of an eyetracking study in the retail sector

Karsten Zischner
Anhalt University of Applied Sciences - Department of Informatics and Languages
Lohmannstr. 23, 06366 Koethen, Germany
Email: k.zischner@inf.hs-anhalt.de

***Abstract*—Today, usability testing in the development of software and systems is essential. A stationary usability lab offers many different possibilities in the evaluation of usability, but it reaches its limits in terms of flexibility and the experimental conditions. Mobile usability studies consider consciously outside influences, and these studies require a specially adapted approach to preparation, implementation and evaluation. Using the example of a mobile eye tracking study the difficulties and the opportunities of mobile testing are considered.**

***Index Terms*—eyetracking, mobile eyetracking, mobile testing, mobile usability, usability, user experience**

I. Introduction

WORLDWIDE we can observe that the field of usability is paid more attention to when developing software and systems. According to DIN EN ISO 9241-11, usability is ”the extent in which a product can be used by specific users in a specific context to achieve certain goals effectively, efficiently and satisfactorily.”[1] In addition to the target and the relevant context of use the user is clearly put in the centre. The technical options should be geared towards the needs of the user and not vice versa. In the middle of the 1980s this requirement was formulated by Donald Norman and Jakob Nielsen due to massive problems in dealing with technical artefacts. Finally, usability is also an indicator for user acceptance as an effective and easy-to-use product will cause no frustration to the user rather a positive user experience will be created. Thats why emotions play an important role. This is known as user experience[2], an expanded concept of usability, especially when the emotional aspects are considered in the context of use. It can be seen that studies on movements of the user focusing on his behaviour can be conducted in many areas of human life. A laboratory is not always the best place for these observations. Denoting the user for example as customer and the technical artifact is a sales area, then give yourself more important applications for usability studies.your conference editor concerning acceptable word processor formats for your particular conference.

This is where the mobility comes in, as you are faced with the following question before the study: If you build a sales area in a laboratory environment, or do you test directly where a real customer wants to make a real purchase.

Both approaches have their place and should be discussed prior to an investigation. The mobile testing provides compared to the laboratory testing significant added value: the testing under real conditions. Those conditions, however, in terms of data collection can also be viewed as potential sources of error causing for mobile testing a specially customized approach to preparation, implementation and evaluation.

Laboratory tests are problematic for purely logistical reasons[3]. It is at this point explicitly noted that a laboratory test can complement a mobile test very well. How and in what way this is possible is exemplified here. Usability testing is comprised of various methods for evaluation. The use of a particular method depends heavily on the context. For laboratory tests and for mobile usability studies the eye tracking is an adequate means. Below the example of the mobile eye tracking used for shelf tests in retail is examined. As a measurement tool, the mobile company Tobii eye tracker (Tobii Glasses) is used. The device offers the advantage over similar tools: extreme compactness and lightness. Therefore it is perceived by the subjects to be very handy and not disturbing.

II. Preparation of a mobile test

A. Necessity

Each evaluation of usability requires a preparation phase. In a mobile test this phase is particularly important. A mobile test usually runs always in a different environment. So local

[1] DIN EN ISO 9241 (1997), S.94. Ergonomische Anforderungen fuer Buerotaetigkeiten mit Bildschirmgeraeten. Berlin: Beuth Verlag

[2] vgl. Sarodnik, F., Brau, H., Methoden der Usability Evaluation Wissenschaftliche Grundlagen und praktische Anwendung, 2.Auflage, Bern: Verlag Hans Huber, 2011.,S.22

[3] Think, for example, on transportation costs and to the structural conditions of a laboratory environment.

conditions vary from test to test. This is an important point that needs to be thought through in preparation. Furthermore, a content preparation is not only useful, but as with any usability study also imperative. As the preparatory phase can be designed specifically for a mobile test should be explained in the following section.

B. Problem definition

Mobile usability tests as mentioned in the introduction may facilitate specific mistakes in application as well as special approaches. The location in this case is a shop with an extremely broad range of products. There is a wide range of applications for mobile usability tests and, at the same time, a great variety of possible research issues that can be worked on with this type of test. Normally, in a shop, we look at the purchasing behaviour of clients often with the help of shelf tests[4]4. The shop management mentioned problems with the organic products range, which generate, despite the general trend, little sales in this shop.

C. Content preparation

In order to carry out a mobile usability study, conceptual considerations are necessary to provide as a result a test scenario with a specific problem. For the development of this test scenario, first the weaknesses in the previous processes are revealed. For a study in the retail sector the customer is to be regarded in their shopping process.

To identify a specific target group and to find within this target group members and typical typical customer, personas[5]5 can be created. Prerequisite for the creation of personas are real information as can be obtained through interviews and initial field observations.

The basic structure of a persona can be represented as follows:

- real Name
- realistic photo of the person
- demographic information
- profession
- objectives, expectations, wishes
- preferences, hobbies, dislikes
- quote

If you got some concrete information and found out the typical customer, a review with the client is an important step. The problem notified in II-B - the relatively weak sales of organic products - must be substantiated before a possible investigation. To this end, offers an expanded interview method that looks at the actual purchase process and thereby represents the behavior of the customers in the sales area. This observation and interview method is known as Contextual Inquiry[6].

The result of the substantive preparation of a mobile usability study is to create a test scenario with concrete tasks. The task is derived directly from the problem. If the problem e.g. The low product sales of a particular product group, then the task is to deal with the Why.

First evidence already produced a survey of 54 customers in the market. The results of two contextual questions illustrate the current situation:[7]

Question 1: Do you have already bought organic products in this store?

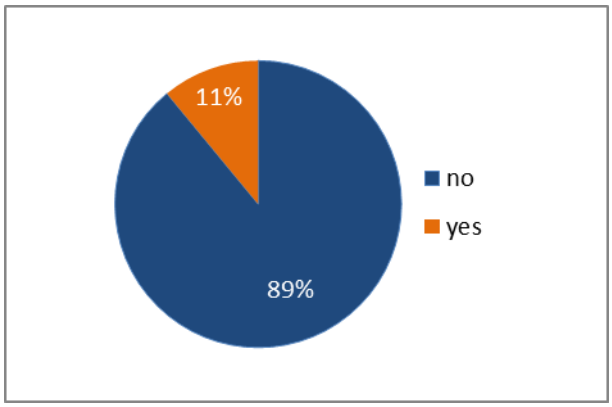


Fig. 1 - Answer to question 1

Question 2: Do you expect the offer of organic products in this store?

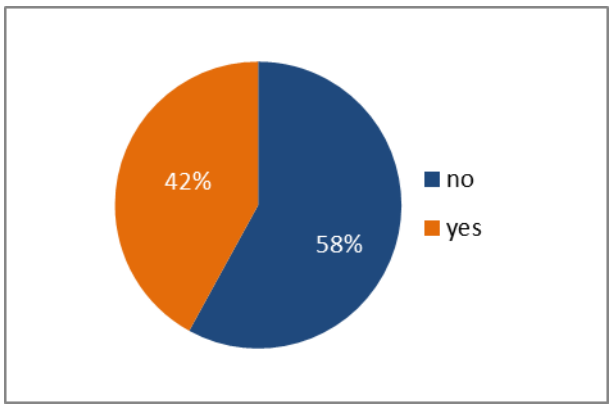


Fig. 2 - Answer to question 2

Nearly 90 percent of respondents in this store never bought organic products (Fig. 1). More than half of respondents do not expect organic products in this retail store (Fig. 2) Even with these two pieces of information it can be illustrate how

[4] more information about shelf tests can be found in the book: Eyetracking im stationaeren Einzelhandel (2005), Dr. Nadine Berghaus, Koeln: Josef Eul Verlag GmbH

[5] vgl. Richter, M., Flueckinger, M.,Usability Engineering kompakt – Benutzbare Software gezielt entwickeln, 2.Auflage,Heidelberg: Spektrum Akademischer Verlag, 2010.,S.28ff

[6] vgl. Richter, M., Flueckinger, M.,Usability Engineering kompakt – Benutzbare Software gezielt entwickeln, 2.Auflage,Heidelberg: Spektrum Akademischer Verlag, 2010.,S.21ff

[7] These questions are part of a survey, which was ahead of the investigation directly addressed in the market to customers.

the supply of organic products currently carried out in this market generally. Therefore a method of investigation is to propose, which may appoint the reasons for this condition. A mobile eye tracking study puts the visual aspect of a purchase process in the focus of attention. Maybe a product is too insignificant? A product can't be found? Draws a product the customers attention on itself? These are examples of typical questions that can be examined with an eye-tracking test.

D. Technical preparation

The focus of the technical preparation of a mobile usability study is the first evaluation of the necessary hardware and software. In the case considered here, the suitability of a specific eye tracking system will be established. The following requirements have to be met:

1) The eyetracker has to be mobile.
2) The eyetracker must be up and running quickly.
3) The eyetracker should not disturb the customer during shopping.

A device that meets these requirements is the mobile eye tracker by the company Tobii, entitled: Tobii Glasses. The name already suggests that it is an eye tracker in the form of glasses. The unit is very light and after a short period of familiarization of the subject hardly noticed. After a test the eye tracker Tobii Glasses is ready for use immediately.

Nevertheless, such a device places special requirements to the test staff. To ensure a smooth testing process, an introduction of the research team, which meant the testers, is absolutely necessary. The calibration of the eyetracker, which is essential to any direct test preparation can be described as particularly challenging. A test is designed to produce usable data, if the calibration performed carefully. The testers need to practice this procedure frequently, thereby to get into a routine. During a field test may not happen in this case as a glitch because that would seem unprofessional on the subject, time-wasting and could be the subject may move to cancel the test.

It is also important at this preparatory stage, to plan the overall technical structure and test examples, since the calibration of the glasses is only a technical component during eye-tracking session.

For efficient data analysis, IR marker[8] are used, during the test a unique ID[9] sent to the glasses. Characterized additional information can be introduced into the receiving material, whereby a positioning of the test person is possible. How does the use of the IR marker affects an automatic analysis of the data is carried out at a later point. The team must practice not only dealing with the eye-tracking glasses, it also needs to know the opportunities and requirements of the complete equipment.

E. Planning the implementation

The planning effort for a field test is much higher than for a test in the laboratory. Therefore the familiarization with the real test environment should not be made until the day of the test. A sales area shall not claim to be an ideal test environment. In order for a mobile eye-tracking study nevertheless obtain usable data, it must be a logistical planning site.

The scheduling plays a very important role. The optimum time of testing was coordinated with the market management[10].

To be seen as a test team to be, any tester must be clearly identifiable as such[11]. It is a special role that the testers belongs during the investigation, and this should also be clearly visible to the outside.

The positioning of the equipment is another important planning step. You need an area for the calibration process, you have to test whether the IR markers to attach the shelf, you have to look at the entire case consistent sales area, the influencing of the test persons by the construction technique has to be avoided. It must be collusion with the staff. If employees are not informed, they are unsafe against the subjects. A mobile eye-tracking test is considered a normal purchase history. These include e.g. the possibility of responding to the employee, if a specific product is not found[12].

The distribution of tasks within the test team must be clearly regulated. The following services have to be provided by the test team for the mobile eye tracking study:

- Approaching potential test persons for the purpose of acquiring
- Calibration of the mobile eye tracker
- Test observation
- Adoption and thanks

The methodological approach and the specific test procedures are described in more detail in the following section.

F. Methodical and chronological sequencing of the test

In order to ensure the most comprehensive answer to the question elaborated, a mobile usability study should include not only the test itself. In the previous part different methods are named, some of these may occur in succession, partly in parallel. Therefore we speak of a mixture of methods. In the case under consideration the combination of methods consists

[8] IR markers (infrared markers) have two modes of operation - when used with a suitable holder the marker is assigned a unique ID, without attachment the marker is used for the calibration of the mobile eye tracker.

[9] ID (identification number) - are in the basic 30 markers + holder and thus 30 unique IDs. The software Tobii Studio can handle with a larger number of IDs.

[10] On weekdays, where goods will be given to the shelves, are ranges of products in the aisles. The normal sales situation is altered. Therefore a test on such a day is unfavorable.

[11] For this study each tester was equipped with a name tag to make the membership of the research team identified.

[12] The employees were in this case explicitly instructed the subjects to be treated as ordinary customers and provide necessary assistance.

questionnaires, eye tracking, interviews and field observations. The questionnaires are part of the test documents. To claim the participants not longer than necessary, the questions must be formulated clearly and allow a rapid response. Answering the questions will take the form of an interview, so the tester has the possibility to minimize the time spent. For the test, a shopping list was prepared by the test team, which has to been processed by the test person. The 5 items on the shopping list were not selected randomly. First there was a study of the so called Rennerliste, which contains information about the sale of the individual product categories. If data has to be collected by questionnaire before the actual test, so these issues should not affect the test participants. If any information is given to the test person, the test sequence will be changed. Initially used questionnaires may not receive any information about where certain items are located. Under certain circumstances it may be appropriate to share a questionnaire and access some information after the test. This approach should also be used for the investigation of the considered application. The questions before the test used for statistical surveys and the assessment of the purchasing behavior of the test participants. After the test, the participant has to answer short questions to test progress. In the final questions to the test person 5 product photos are shown, which were not on the shopping list. The test team developed for this part of the investigation a memory test, which checks the quality of perception of certain products. To collect these information, the making of special, meaningful product photos is necessary[13].

In a usability test in general, and in a mobile test in particular, the creation of equal conditions is very important. This is also called a stereotypical test procedure. It has been shown that a checklist for the performance of the test is very advantageous. Thus, not only have the same sequence of the individual test sequences are adhered to, it is also necessary to ensure that all test subjects received the same information before the test. Even the wording of the explanations to the test persons should be the same as with every test.

When considering the time required, the test team agreed that it is necessary to carry out the tests estimate at least one working day. Per subject, the entire test should not take longer than 15 minutes. This is due to the fact that the subjects will not be acquired in advance of the test, as is customary for a test in the laboratory. The potential subjects are real customers who want to carry their shopping in this market. This daily process should be the basis for the proposed eye tracking test. Since it was suspected that many customers can spend time for the test, the expectations regarding the expected number of subjects were very moderate. The target of the test team was to perform at least five usable tests.

Other considerations were concerned with the question of how to complement a mobile eye tracking test methodology. In the planned test in a sales area it comes out mostly about whether certain products with as little problems can be found. A single product here to look so detailed that you can assess whether the customer focuses more on the product label or on the product image, can afford a mobile eye-tracking test barely. Here, the test team developed as an extended test method a special stationary eye-tracking study that can provide additional information. For this test, the product photos that were used in the memory test are reused. The extended methodology clearly shows that mobile testing can be usefully complemented by stationary tests[14].

III. Carrying out the test

A. Test setup in the market

The test setup in the market must be made prior to the sale operation. For setting up the test area for a mobile task based eyetracking study the following tasks have to be executed:

Install and set up the IR markers: IR markers are optional elements for an eye-tracking test. They allow fast automated analysis of the data with a **GazePlot** or a **Heatmap**. The IR-marker must be placed on the shelf to beconsidered.

Creation of a calibration area: Before each test, calibration of the Tobii Glasses is necessary. For this purpose, a sufficiently large, monochrome, uniform surface is required, that must be prepared and installed at a suitable location[15]. The subject needs to set up the eye-tracking glasses at a distance are of exactly one meter in front of this area. The tester must be equipped with an IR marker constitute a virtual raster, which look at the test person. The actual calibration area could be built in a low-traffic area of the market.

Provide a desktop environment for completing the questionnaire: As the response to the questionnaire will be moderated by a member of the test team, the calibration site used for this task. It is important to ensure that test takers receive no insight into the test data and running tests of other test takers. It must be ensured that a test has been completed, before another new participant enters the test environment.

Thanks to test persons: The active participation in such a study is not self-evident. Customers who agree to participate in the study will receive a small gift for their help[16].

Ensure ready to use test equipment: The entire experimental technique works with batteries. In a field test, sometimes it can be a problem when batteries need recharging. Therefore charging options has to be provided before the test starts. This is particularly important for the IR markers used for calibration and for the recording device. The battery capacity of the recorder should not be completely exhausted, because you cannot judge exactly how many minutes the device must record before the test. Overall, the recording

[13] For this test, the product photos specially tailored to use them in subsequent laboratory tests.

[14] The valuable information is retained about a mobile application scenario, but they are enriched with additional facts from the laboratory.

[15] The required surface was realized by the research team with the help of two flipcharts.

[16] It is recommended to provide these little gifts visible before the test, to create an additional incentive for participation.

device of the Tobii Glasses can record about 70 minutes continuously with one battery charge. By the required calibration prior to each test, in reality this time is not achieved. The charge of the IR marker on the shelf is not critical. These are usually active only when the subject with eye tracking glasses is nearby.

Preparation of the methodical process: For a rapid test procedure, it is necessary to provide all the test papers before the examination. The workflow for the test has been already discussed here but it needs to be went through step by step in the real environment. A dry run is recommended.

B. Chronological sequence of the test

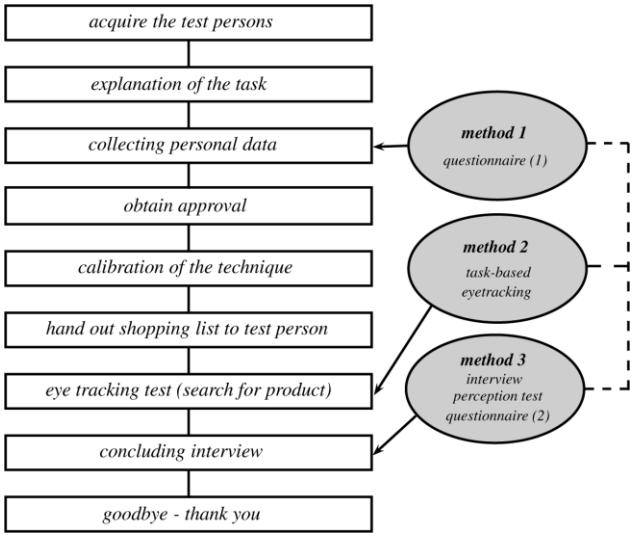


Fig. 3 - Chronological sequence of the test

Considered for the mobile eye-tracking test a chronology was developed, which forms the basis for the collection of comparable data. A deviation from this order may affect the reliability of the data collected and thus corrupt the negative results of the test. Fig. 3 brings the sequence of tests with the use of usability methods in relation.

IV. REMACHINING THE TESTS

A. Archiving and import of the test data

The test data are recorded on SD card[17] during the test. Each recording receivs a unique session ID. The SD card must be read and imported into the special software Tobii Studio[18] for the analysis.

B. How to Create a PostScript File

Using the Tobii Studio software, the next step must determine the suitability of the individual recordings. This may rely very important on observations made during the test. Therefore, it is recommended that at least a part of the tester is also involved in the evaluation of the data. The operations at the mobile testing are layered and complex that the individual test persons can be assessed only by knowing the complete test procedure. For a comprehensive documentation that brings a separate evaluation team to the same level of knowledge as the test team, remains no time during the field test experience.

A mobile eye tracking study first requires the conscientious performance of the calibration of the eyetracker. The changing of the position of the eye tracker during the test by the test person, caused erroneous data. These data are not suitable for subsequent analysis[19].

C. Data treatment using parallax correction

Depending on the eye tracker, there are factors that can be corrected automatically. The eye tracker of the company Tobii, which was used for this study, allowed to correct the so-called parallax error using the Tobii Studio software. This occurs due to the fact that the calibration of the eye tracker is a fixed distance of one meter. During the test process, consider the subject specific objects from a distance or in a very short range. So there is a displacement of the actual subject region. According to data analysis, a test person has not seen a particular product, even though the subject has taken this product off the shelf. This shift can be achieved out by using the parallax correction within the software Tobii Studio.

D. Automatically generated evaluations

GazePlot[20] and Heatmap[21] are visualizations that can be created using the Tobii Studio software. There are eye movements (fixations[22] and saccades[23]) or represented clusters of fixations. At a mobile eye tracking study specific areas are considered of particular interest. For a shelf test, this interesting area can be a full rack or a certain part of a shelf. Since each product follows a specific sequence, which implies by the test person itself, the data, collected for an automated analysis cannot be directly compared. This is an important difference between a mobile and a stationary eyetracking study. For a mobile test, test subjects can approach to be examined the shelf from different directions, or they can look at the shelf from a small or a large distance. To make the collected data more comparable tests, we made a photo prior to the examination (see Fig. 4) from the observed shelf or from a part of the shelf on. For this recording, the IR markers have been placed on the shelf. Since each marker has a unique ID, must be an accurate identification of the marker within the

[17] SD Card - SD Memory Card (Secure Digital Memory Card) is a digital storage medium. For the recording device of the mobile Eyetrackes Tobii Glasses an SD card with a maximum capacity of 4GB (4 Gigabyte) is used. In order for the recording of about an hour of video, including all metadata.

[18] Tobii Studio version 2.3 (and above) can handle data with the Tobii Glasses.

[19] For more information about error detection and error handling can be found under item IV-F.

[20] A GazePlot represents eye movements as a chronological sequence of fixations and saccades

[21] A Heatmap identifies areas that were considered long or often.

[22] A fixation within a view profile is shown as a point range.

[23] Two fixations are connected within a view course by a saccade. This is no information acquisition.

Tobii Studio software. Therefore it is important, the use of the markers to be documented in the device exactly, i.e. to note the ID number of each marker.

Fig. 4 - wide-angle lens recording of the shelf

By using the Tobii Glasses itself for making the photos (see Fig. 5), this marking is done automatically[24].

Fig. 5 - Photographic recording of the shelf with Tobii Glasses

Automatically generated evaluations based on static AOI[25] are very attractive when fast initial results to be made visible. A critical examination of the data obtained is essential. It has been shown that under certain circumstances there is no detection of the eye movement data. This is e.g. due to the fact that for a precise positioning of the subject must always include at least 4 IR markers in sight. At close range, i.e. if the customer is close approaches to the shelf for taking a product, this may lead to problems.

[24] Due to the local conditions in a sales area a photograph with an external camera with wide-angle lens is recommended. When comparing the figures 4 and 5 perspective distortion when shooting with the Tobii Glasses can be seen

[25] A AOI (area of interest) is an individually defined area which is regarded as a test of the subject, and for which an eye movement analysis is of great importance. A static AOI is featured on a photo on which the eye movements of all tests can be aggregated.

E. Work with dynamic Areas Of Interest

Besides the use of an static image to bundle the eye movements of all tests, as described in section IV-D, so-called dynamic Areas Of Interest (dynamic AOI[26]) are used for evaluation. The advantage of this approach is the independence of the IR markers. Data are recorded at close range. The creation of dynamic Areas Of Interest is very complicated, but should be used if the AOI based on static, automatically generated evaluations on the basis of test observations are considered to be faulty.

F. Error estimation

While eye tracking studies in stationary automatic evaluations are based on static AOI, can consider to be quite reliable, the application is considered critical in field trials. Local conditions play a major role, because a store shelf is not designed for the inclusion of IR markers. You have to make compromises which sometimes have incorrect or missing data as a result. Alternatively, the more complicated method of application of dynamic AOI can be used. Both methods were described in the sections IV-D and IV-E.

The data derived from the two methods are very comprehensive and should not be considered complete at this point. For an error estimation example, the value Total Visit Duration[27] chosen for two specific products (see Fig. 6). The differences are sometimes considerably and give the reason for the need for the critical analysis of the data determined automatically.

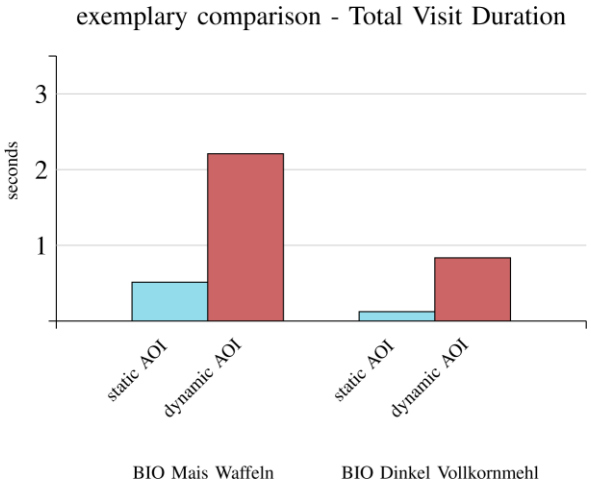


Fig. 6 - error estimation

G. Evaluation of data

As shown in section II-F the use of a mix of methods is an

[26] Dynamic AOI and static AOI in terms of the creation. Dynamic AOI on the changing media (videos) are created. creating dynamic AOI is very time consuming because the size and shape of the AOI must constantly change.

[27] The average Total Visit Duration is calculated from the duration of individual fixations in a given AOI divided by the number of tests

adequate means in a mobile usability study for a comprehensive analysis of usability problems. For the analysis of the data from this approach consequences resulting directly. A methodological analysis is applicable to a first rough estimate, for a deeper analysis, a method-disciplinary approach can be selected. The team needs to evaluate this specific data and discover relationships. Much is possible, not everything makes sense. An example at this point is to the buying experience, which was determined during the interview before the test, and the time for finding a certain product are brought into relation. The following questions were used to assess the buying experience:

Question 1: How often do you shop in this store?

Question 2: Are you aware of discount offers at this store?

During the eyetracking study of each test subject the times for locating specific products were determined. Bringing these data together with the data on the buying experience, the following picture emerges:

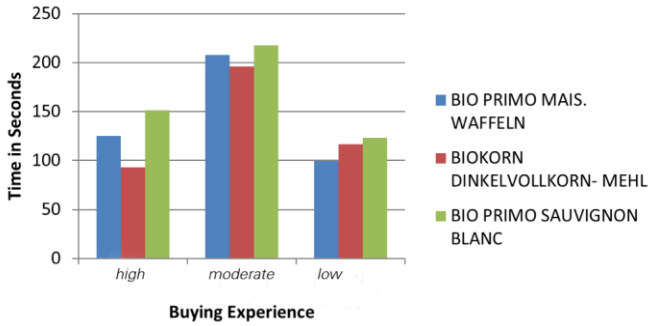


Fig. 7 - buying experience and time to find the products

It can be seen that the buying experience in the search for organic products only plays a minor role. Even experienced buyers (regular customers) have trouble finding the products.

H. Methodological supplement of a mobile eye tracking test

As discussed in Section II-F discussed, certain questions will be answered by additional tests in a usability lab. For these methodological supplement of a mobile eye tracking tests, a steady eye tracker was used. Six test persons were given the task to see product images as a slide show.

What is a customer oriented when viewing a specific product?

In answering that question on any product image, three AOI were created and grouped accordingly:

- AOI Group 1 - product photo
- AOI Group 2 - Bio-Logo
- AOI Group 3 – scripture

Fig. 8 - example - with and without AOI

Looking at the time to first fixation[28] it can determine that the Logo acts as an eye catcher. The time to the first fixation in this area is the shortest. The value of total fixation duration (including zeros)[29] for the scripture area is the greatest. This allows conclusions to the general reception of information. For a detailed consideration other values are to be determined, but at this point not to be displayed completely. This analysis delivered the result that the section of the scriptureis is the most relevant area for recognition and remembrance.

V. Conclusion

A. Unforeseen and unforeseeable difficulties

When planning a mobile usability study, the test team must discuss any difficulties, which can be divided into two groups:

a) from a technical perspective: The risk of technical problems is given both in the laboratory and in a study in the field. It is critical, however to evaluate the case of mobile usability studies in the field. The test team has to make a sufficient number of storage devices available and provide for the recharge of the battery. Regarding the calibration it has been suggested prior to the examination that the lighting in the market could complicate the calibration procedure. The time required for calibration could increase, the inaccuracy may be higher, which can potentially lead to abortion of the test. In practice, these assumptions are partially confirmed. The lighting conditions were not optimal, nethertheless an abortion of the test was not necessary.

b) regarding the test persons: Before a mobile usability study there is no guarantee that a sufficient number of subjects is available. The subjects to be an investigation in the field is not usually invited. They are accessed spontaneously, and to refuse to participate. Presumptions that too few customers

[28] is the time a test person first looked on the AOI

[29] calculates the average total fixation duration (including zeros) by the duration of individual fixations in a given AOI group divided by the number of recordings.

could be aquired for a test were not confirmed in practice. The great popularity led however to a technical problem. It had considerably more calibrations are carried out as planned, which claimed the battery of calibration markers than.

The potential problem areas have been identified before the test though, concrete difficulties were not entirely predictable.

B. How to Create a PostScript File

Finally it can be stated that the positive aspects outweigh the considered here mobile usability study. Even some of the expectations were surpassed the test. Particularly positive subjects acquisition is emphasized.

It has also shown that the preparation and evaluation phase are particularly challenging. Each task has to be reconsidered for a test in the field. We know the tools, but they must be put together for each viewing again and again.

REFERENCES

[1] Sarodnik, F., Brau, H., Methoden der Usability Evaluation Wissenschaftliche Grundlagen und praktische Anwendung, 2.Auflage Bern: Verlag Hans Huber, 2011.

[2] Berghaus, N., Eye-Tracking im stationaeren Einzelhandel - Eine empirische Analyse der Wahrnehmung von Kunden am Point of Purchase, 1.Auflage Koeln: Josef Eul Verlag, 2005.

[3] Richter, M., Flueckinger, M., Usability Engineering kompakt – Benutzbare Software gezielt entwickeln, 2.Auflage Heidelberg: Spektrum Akademischer Verlag, 2010.

[4] Nielsen, J., Usability Engineering, London: AP Professional Ltd., 1993.

[5] Vrizi, R.A., Refining the test phase of usability evaluation: how many subjects is enough? Human Factors, 1992.

[6] DIN EN ISO 9241, Ergonomische Anforderungen fuer Buerotaetigkeiten mit Bildschirmgeraeten.,S.94. Berlin: Beuth Verlag, 1997.

Influence of Jitter on Reliable Multicast Data Transmission Rate in Terms of CDN Networks

Aleksandr Bakharev, Eduard Siemens
Anhalt University of Applied Sciences - Department of Electrical, Mechanical and Industrial Engineering
Bernburger Str. 57, 06366 Koethen, Germany
E-mail: a.bakharev@emw.hs-anhalt.de , e.siemens@emw.hs-anhalt.de

***Abstract*— This paper devotes to evaluation of performance bottlenecks and algorithm deficiencies in the area of contemporary reliable multicast networking. Hereby, the impact of packet delay jitter on the end-to-end performance of multicast IP data transport is investigated. A series of tests with two most significant open-source implementations of reliable multicast is performed and analyzed. These are: *UDP-based File Transfer Protocol* (UFTP) and *NACK-oriented Reliable multicast* (NORM). Tests were targeted to simulate scenario of content distribution in WAN – sized Content Delivery Networks (CDN). Then, results were grouped and averaged, by round trip time and packet losses. This enabled us to see jitter influence independently on round trip time (RTT) and packet loss rates. Revealed jitter influence for different network conditions. Confirmed, that appearance of even small jitter causes significant data rate reduction.**

Keywords: jitter;reliable multicast; CDN

I. INTRODUCTION

A. Terms definition

We are defining the term "*jitter*" or "*delay variation*" in accordance to RFC-3393 [1]. Assume that "jitter" or "delay variation" in networking is a difference of one way network delay values for two consecutive packets with respect to some reference value. Practically, it means that: jitter value 10 ms corresponds to fact that RTT in network will vary in range of about 20 ms from certain value. Practical example shown below:

```
   64 bytes from 192.168.11.1: icmp_seq=15 ttl=64
time=124 ms
   64 bytes from 192.168.11.1: icmp_seq=16 ttl=64
time=150 ms
   64 bytes from 192.168.11.1: icmp_seq=17 ttl=64
time=127 ms
   64 bytes from 192.168.11.1: icmp_seq=18 ttl=64
time=132 ms
   64 bytes from 192.168.11.1: icmp_seq=19 ttl=64
time=106 ms
   64 bytes from 192.168.11.1: icmp_seq=20 ttl=64
time=124 ms
```

So, jitter affects inter-arrival gap and consequently leads to two effects, such as *clumping* (inter-arrival gap decrease), and *dispersion* (inter-arrival gap increase).

Main evaluation metric for the transmission performance is the achieved *data rate*. When calculating data rate, assume that data rate is the relation of amount of transmitted data over passed time:

$$R = \frac{A}{T}$$

Where R – data rate in Mbits/s, A – amount of transmitted bits, T – transmission duration in seconds.

It was also stated that results will be *averaged* by two parameters (packet loss rate and RTT). It means that we will have two additional specific types of achieved data rates:

$$R_{RTT\ independent} = \frac{R_{RTT_1} + R_{RTT_2} + \cdots + R_{RTT_i}}{i}$$

$$R_{losses\ independent} = \frac{R_{loss_1} + R_{loss_2} + \cdots + R_{loss_i}}{i}$$

By intention, calculation of e.g. $R_{losses\ independent}$ for different RTT values, we can see how jitter affects network performance without influence of losses on result. This metric is to be introduced due to probability of too high sensitiveness of protocols to RTT or losses. Averaging of results could minimize these effects and give more "pure" results.

B. Background

Jitter appears in any packet switched network and can significantly affect data transmission quality. Main issue is that jitter causes breaking of timing logics in transport protocols. If variation of inter-arrival interval exceeds some certain range, a situation occurs, that packets are considered dropped, though they are arriving just late. Consequently, receiver could be just not ready to accept packet, which is arrived not in time.

In real networks, significant jitter usually appears in cases, when OWD in network is relative high (practically it means long distance between network units). From this fact, occurs the first main reason of jitter – unexpected delays at intermediate devices (routers, switches, e.g.). This could be caused by buffer or performance issues at the intermediate device. In fact, unpredictable delay is the main part of jitter value.

However, jitter is caused not only by queuing within network nodes, but at the sender's side as well. Main source-related reasons of jitter at the data sender are sender's timer

and scheduling issues. When using non-precise timers, we could not be sure about initial inter-packet gap and this uncertainty transforms in jitter effect at the receiver's side.

Approaches to minimize jitter influence coming from its reasons. Generally speaking, only what could be done are:

Implementation of *jitter buffers*, like in *Real Time Protocol* (RTP) [2] stacks of VoIP and Video-over IP applications. This is just an additional buffer, which stores certain amount of data packets on arrival and timely transmit them to the receiver, thereby, emulating correct inter-arrival gap for the client.

Adding additional interactivity to communication protocol, as it was done by *TCP-interactive* protocol [3]. Practically it means better analysis of network within data transmission and dynamic adjustment of transmission parameters.

The goal of this paper is to analyze the impact of jitter on actual data rate of reliable IP multicast transport protocols. Based on previous experience with evaluation of the data transport number contemporary reliable multicast approaches [4] and evaluation of multi-gigabit reliable transport protocols [5] we have expected that jitter influence would be not so significant in reliable multicast networks, at least in frame of our experiment, due to relative low data rates.

II. EXPERIMENT MAP

As the test scenario, a delivery of 4.1 Gbytes file among 3 receivers, by means of reliable multicast data transmission has been chosen. Such a scenario assumes quite easy topology and enables to observe more "pure" dependencies in the network due to minimizing of intermediate devices. One more reason for this topology is that our initial goal is to emulate distribution of heavy content in CDN network. Such a scenario often does not assume hundreds of recipients. We are dealing here with only three recipients and it enables us to observe jitter effects in not overloaded network. By means of such network conditions we got rid of additional network issues, such as congestions, receive/send buffers overflow and etc. Network impairements (delay, losses jitter) are managed by Netropy 10G network emulator, manufactored by Apposite Technologies [6]. The network has been set up in 10G laboratory of Anhalt University of Applied Research in Koethen (Germany). Topology is presented at Figure 1.

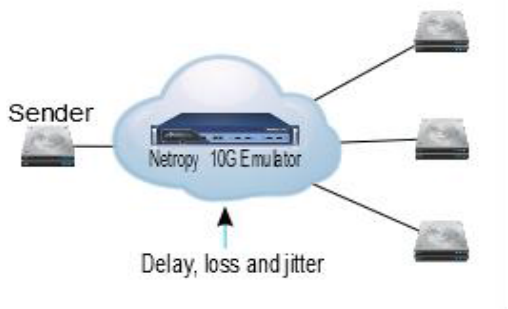


Figure 1. Test network setup

All participants of multicast data transmission session are Linux-based servers. Sender has Centos 6 installed at the Intel i7 machine, while all clients are running Open Suse 11.4 Intel Atom machines. All systems are running 64x editions of OSes. Network impairments have been set to the following values:

RTT: 60, 120 (ms)

Packet loss rate: 0.0, 0.1, 0.7 (%)

Jitter: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 (ms)

Protocols under consideration are *NORM* [7] and *UFTP* [8]. Though in the performance analysis paper [4] openPGM [9] was also taken into account, however initial performance tests have revealed inability to use this protocol for Big Data transmission.

III. PROTOCOLS

A. NACK oriented reliable multicast (NORM)

The *NORM* protocol was defined within RFC 5740 [7] in year 2009. The source code of a reference implementation of NORM is maintained by the Naval Research Laboratory [8]. It delivers along with the transport protocol a ready-to-use application, which can be compiled from available C source code on Linux. The NORM application, offers features like TCP-friendly congestion control which provides fair sharing of available bandwidth between multiple data streams. NORM uses selective NACKs to provide reliability. NORM can also be used in conjunction with FEC, which is actually an on-demand feature.

B. UDP-based File Transfer Protocol (UFTP)

UFTP is a reliable multicast protocol as well as a corresponding end-user application and can be considered as a successor of *Starburst Multicast FTP* (*MFTP)* [10] proposed in 2004 and offers reliable multicast file transfer by means of typical UDP transport. The protocol is currently in use in production of the Wall Street Journal to send WSJ pages over satellite to their remote printing plants [11].

UFTP proposes a specific scheme of data transmission organization. First of all, the protocol decides how to divide an input data set. It is going to be split into blocks whereby one block is always sent within one UDP packet, while blocks, in turn are logically grouped into sections. Divided into blocks and sections, the sender just sends a section to a multicast group. As soon as a transmission of a section is finished, the sender requests current status of received data from each multicast receiver and gets a batch of packets which contain a list of missed packets at the site of each recipient. On reception of all NACKs, missed blocks are retransmitted in the unicast way. The sender will begin to transmit a new section only after the reception of all blocks of the previous section at each recipient in the multicast group.

IV. RESULTS

Previous results (4) showcased that UFTP transfers data much faster than NORM and this fact enables us to observe jitter influence at very different data rates. The result of first run is presented in Figure 2. At the first glance is seen that data rate decrease, caused by appearance of the minimal jitter, is the most significant one. It is going to be referred to the phenomenon of *instant data rate decrease* (firstly

introduced in this paper) at the moment when jitter appears – delay variation of 1 ms. This effect is especially visible at plot, corresponding to UFTP. Without jitter, it reaches data rate of more than 200 Mbits/s, but even with 1 ms of delay variation, data rate degrade till 150-175 Mbits/s and afterwards we observed quite low decrease – till 140 Mbits/s at 9 ms of delay variation. The same tendency can be observed with NORM, but it decreases more drastically and at low data rates.

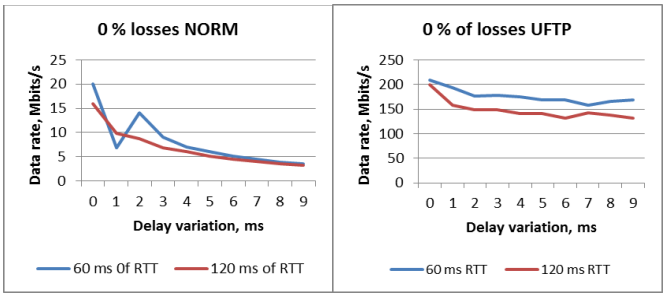


Figure 2. Data rate on delay variation dependency fot link with low loss

Figure 3 extends this exploration. Analysis of Figure 3 points to the fact, that with decreasing of data rate (caused by high delays and losses) data rates become more "stable". We do not see bursts (particularly a significant decrease of data rate). It means that jitter influence decreases with data rate decrease due to other impairments. This is contrastly visible at plot of 0.7 % of losses of the NORM protocol. NORM is pretty sensitive to network impairments, and 120 ms of RTT in conjunction with 0.7 % of losses gives high decrease of data rate (lower than 1 Mbit/s). And exactly at this level, it is obvious that jitter influence almost dissapeared. Table 1 and Table 2 demonstrate some numbers which describe real value of data rate reduction and confirm this statement. For losses-free link there is a reduction by the factor higher than 4. While in high-losses link same parameter is not more than 10% from initial value. Graphicaly, this fact presented at Figure 4.

TABLE I. DATA RATE VALUES, MBITS/S. FOR NORM AT 0.7 % LOSSES LINK WITH 120 MS OF RTT

RTT/ Jitter	0	1	2	3	4	5	6	7	8	9
120 ms	0,80	0,81	0,80	0,82	0,78	0,75	0,74	0,71	0,70	0,70

TABLE II. DATA RATE VALUES, MBITS/S. FOR NORM AT 0.0 % LOSSES LINK WITH 120 MS OF RTT

RTT/ Jitter	0	1	2	3	4	5	6	7	8	9
120 ms	16,00	9,76	8,70	6,84	6,02	5,10	4,52	4,08	3,56	3,25

The same test of UFTP protocol gives somewhat different results. As seen in Figure 5, one sees that all three lines of different packet loss rates are almost parallel. It means that influence of losses on data rate on jitter dependency is minimal for UFTP. Even in links with relatively high losses observe phenomenon of *instant data rate decrease*, mentioned above in the current section. Later on, at Figure 6 slow decrease of data rate take place. In case of UFTP, it was revealed that both, RTT and losses do not significantly influence the way, how jitter affects data transmission rate. However in case of NORM, the absolute value of RTT plays a role, but only on jitter values up to 4 ms. On higher jitter, lines go almost parallels. This dependency is demonstrated in Figure 6.

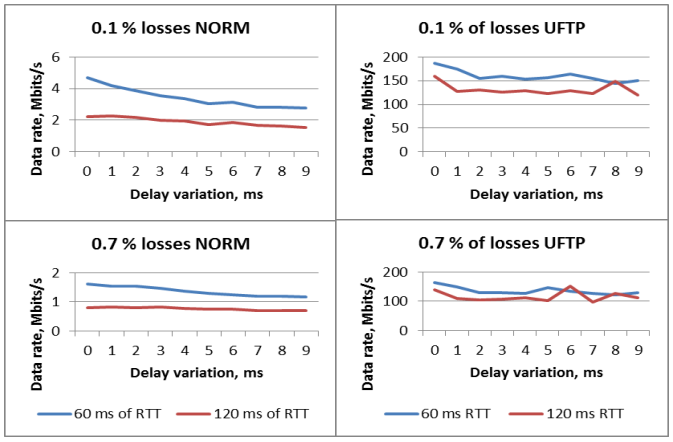


Figure 3. Data rate on delay variation dependency for link with different losses

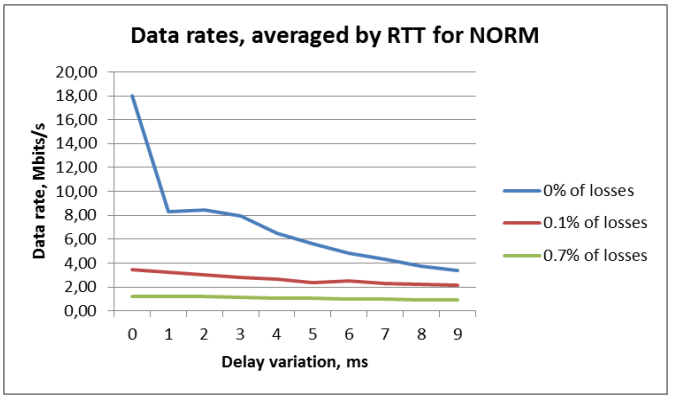


Figure 4. Data rate on delay variation dependency for link with different losses (with averaged value of RTT)

V. CONCLUSION

The reason of *instant data rate decrease* phenomenon is not certainly clear here, however obviously looks like its reasons are in the algorithms of protocols. The fact, that this phenomenon could be observed only at the upper data rate limit of both protocols points to the idea that no one of considered protocols is able to manage jitter issues properly at high data rates. Understanding of this phenomenon going to be discovered in future works of our team.

UFTP behavior looks very stable in any case and is apparently caused by very effective data transmission scheme. Contemporarily, this is the only protocol, which

uses consolidated NACKs [12] in a bit concealed manner - division of input data by sections and blocks - within data transmission, and all our tests showcased its efficiency. Providing of reliability is very resource-consuming process and UFTP manages it very effectively. Consolidated NACKs minimizes amount of active NACKs in the network and lead to the provision of reliability in the most effective manner. This is especially crucial in multicast sessions, when sender has to deal with multiple (of even thousands) of receivers. In such scenario, the price of each NACK is very important.

NORM protocol uses the very common NACK-based retransmission scheme and it leads to relative low data rates, while amount of NACKs at the network is much more than in case of UFTP (due to test setup restrictions, we can't give certain numbers here, but this issue has to be covered by our future works). Processing of each NACK consumes more and more resources with increasing transfer data rates. And practically it makes the protocol very sensitive to any kind any network impairments.

A pretty new finding within this research work is the fact that jitter mostly affects data transmission rate only at high data rates. All test runs confirm that jitter influence at low data rates could be even neglected for both protocols. It could be explained by the fact that at low data rates, timing issues, such as changing inter-arrival gap do not affect receiver significantly. The packet rate is also lower and it enables both protocols to deal with jitter in a very good manner.

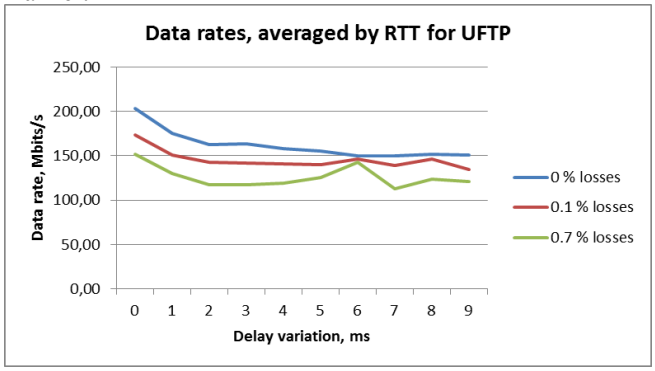


Figure 5. Data rate on delay variation dependency for link with different losses (with averaged value of RTT)

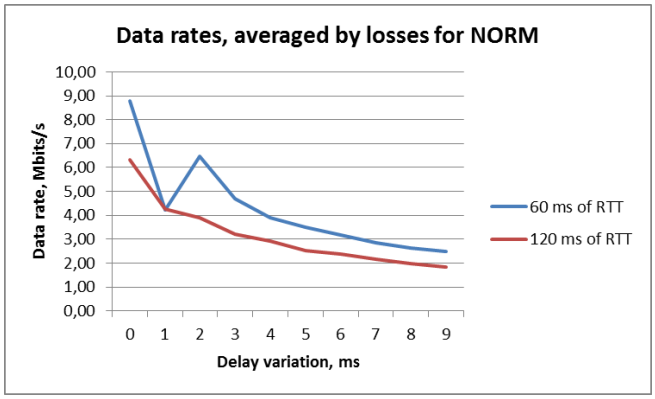


Figure 6. Data rate on delay variation dependency for link with different RTTs (with averaged value of losses)

REFERENCES

[1] C. Demichelis and P. Chimento, "IP Packet Delay Variation Metric for IP Performance Metrics", RFC 3393, 2002.

[2] H. Schulzrinne, S. Casne, and R. Frederick, "RTP: A Transport Protocol for Real-Time Applications," RFC-3550, 2003.

[3] Javed I. Khan and Raid Y. "Jitter and Delay Reduction for Time Sensitive Elastic Traffic for TCP-interactive based World Wide Video Streaming over ABone", Proceedings of the 12th IEEE International Conference on Computer Communications and Networks, ICCCN 2003, Dallas, Texas, October 2003, pp.311-318.

[4] E.Siemens and A. Bakharev, "Evaluation of Reliable Multicast Implementations with Proposed Adaptation for Delivery of Big Data in Wide Area Networks", Accepted for publishing in the proceedings of ICNS 2013, March, Lisbon, 2013.

[5] E.Siemens and D. Kachan, "Comparison of contemporary solutions for high speed data transport on WAN connections". Accepted for publishing in the proceedings of ICNS 2013, Lisbon, 2013.

[6] Apposite Technologies home page. http://www.apposite-tech.com. [retrieved: March 2013]

[7] B. Adamson and C. Bormann, "NACK-Oriented Reliable Multicast (NORM) Transport Protocol", RFC-5740, 2009..

[8] Naval Research Laboratory. http://www.nrl.navy.mil/. [retrieved: March 2013]

[9] T. Speakman, J. Crowcroft, and J. Gemmell, "PGM Reliable Transport Protocol Specification", RFC-3208, 2001.

[10] Internet-draft, StarBurst Multicast File Transfer Protocol (MFTP) Specification. http://tools.ietf.org/html/draft-miller-mftp-spec-02.[retrieved: March 2013]

[11] UFTP - Encrypted UDP based FTP with multicast. http://www.tcnj.edu/~bush/uftp.html. [retrieved: March 2013]

[12] R. Kermode and L. Vicisano, "Author Guidelines for Reliable Multicast Transport (RMT) Building Blocks and Protocol Instantiation documents", RFC-3269. Vicisano, R. Kermode and L. 2002.

A Model of Video Traffic Transmission in IP Networks

Elena Kokoreva, Anna Belezekova
Siberian State University Of Telecommunications And Information Sciences - Faculty of mobile communications and multimedia
Kyrova Str. 86, 630102 Novosibirsk, Russia
E-mail: elen.vic@gmail.com, anna-belezekova@mail.ru

Abstract — Telecommunications and network technology is now the driving force that ensures continued progress of world civilization. Design of new and expansion of existing network infrastructures requires improving the quality of service (QoS). Modeling probabilistic and time characteristics of telecommunication systems is an integral part of modern algorithms of administration of quality of service. At present, for the assessment of quality parameters except simulation models analytical models in the form of systems and queuing networks are widely used. Because of the limited mathematical tools of models of these classes the corresponding parameter estimation of parameters of quality of service are inadequate by definition. Especially concerning the models of telecommunication systems with packet transmission of multimedia real-time traffic.

Keywords: telecommunication systems; TCP/IP; mathematical simulation; queuing system

I. INTRODUCTION

For the transmission of multimedia information in real-time IP technology is widely used. It is known that IP does not provide data transfer with guaranteed parameters of quality of service, however, modern communication services, such as Video on Demand (VoD), place high demands on parameters such as the one-way delay and jitter.

In connection with the abovementioned, the task of the evaluation, and control of parameters of quality of transmission of multimedia data under real load conditions is actual. To solve this, it is necessary to develop the complex of mathematical models that provide assessment of major parameters of quality: one-way delay, the probability of delivery, jitter. This problem is the subject of a number of works of Russian and foreign researchers. So, for example in [1] the impact of the IP network on the quality of the video transmission stream is tested.

Mathematical modeling is one of the main research methods of telecommunications systems. The mathematical model should provide: adequate mapping of various network processes, assessment with the required accuracy the parameters of these processes and the interpretation of modeling results and effective use of computing resources, the use of the model in real time. The class of mathematical models includes discrete or continuous dynamical systems, Markov processes, systems and queuing networks, graphs, hyper graphs, hyper-networks, Petri nets, statistical models, etc.

To evaluate the performance of video transmission traffic in a packet-switched networks often the method of queuing theory and Markov models [2, 3] is used. In [4], the results of performance simulation of the MAC protocol TDMA in wireless networks for a variety of multimedia applications are discussed. Publications in [5-11] are devoted to the application of Markov processes and queuing systems for the analysis of the transmission of multimedia traffic in networks with different architectures (GSM, CDMA, ATM, etc.).

The models based on the use of the apparatus of inhomogeneous closed queuing networks have showed their effectiveness in solving the problem of estimation of parameters of the quality of service in communication systems. Good properties of the method are result of to the fact that the processes of statistical multiplexing most adequately described by the process of queuing, and all possible combinations of the transmitted traffics delays are well described by the topology of these networks, as well as the development of the theory queuing networks and of effective computational methods for calculating them.

II. DESCRIPTION OF THE TELECOMMUNICATIONS SYSTEM

A simulative modeling was carried out for the system shown in Figure 1. Here you can see a telecommunications system, which provides the clients with real-time information service – e.g. video on demand (VoD).

A typical communications system that provides this service consists of the following components:

- a video server connected to the switch IP network through a Gigabit Ethernet link;

-a client workstation, which is interconnected with its own switch to a Fast Ethernet link;

- multiple switches provide transfer of traffic from the server to the subscriber and back.

Video stream that is transmitted to the client workstation is encoded in MPEG-4 [12]. For transmission at the transport layer protocols RTCP and TCP are used, on a network - protocol IPv4.

In this model traffic going to the switch from/to other clients, is a background load, which can significantly affect the delay and jitter.

For the mathematical description of the system (Figure 1) it is necessary to specify three classes of parameters:

- parameters of information load;
- parameters of telecommunications equipment;
- software settings.

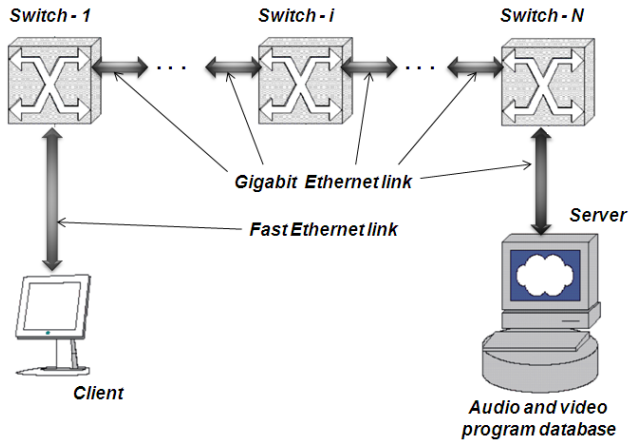


Figure 1. The structure of IP networks for video traffic

III. DESCRIPTION OF THE ANALYTICAL MODEL

For the evaluation of quality parameters of the transmission of multimedia data, including jitter of one-way delay video stream, an analytical model was developed in the class of inhomogeneous of closed queuing networks. In the proposed network model of service, its classes of requirements display the corresponding information flows that occur in the transmission media stream. The nodes of this model display elementary delays that undergone by the elements of these flows in the various components of the telecommunication network, when passing the route of delivery.

The proposed queuing networks (Figure 2) take into account the composition and the topology of the telecommunication network, technology used, as well as the parameters of information load, defined by the information service video on demand. Network service model consists of L nodes and K classes of requirements, the number of which is determined by the structure of the traffic by the size of different types of frames, by the number of intermediate switches, by the parameters of communication lines, as well as by the background load on the transport subsystem.

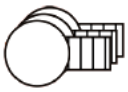
The developed model takes into account hardware, software, and information parameters of the simulated system. It allows us to estimate one-way delay jitter, as well as obtain a wide set of probabilistic-time characteristics, namely: the coefficient of utilization of various types of equipment; the delay in the transmission of video frames of each type; on the average number of IP packages for different purposes in each communication site, network bandwidth, etc.

A. *The components of the network model of service*

Nodes of network model of service are used to display various types of transmission delays and processing elements of the information flows in telecommunications equipment of modeled of system. In Figure. 2 to display the nodes of network of service, different types of services in use, the following notations are used:

- Used to refer to the node of network of service, which is the queuing system, type G / G / 1 / ∞, which uses a service discipline FCFS (First Come First Served, service requirements in order of arrival);

- Used to refer to the node of network of service, which is the queuing system of type G / G / ∞ / ∞, which uses a service discipline IS (Infinity Server, service requirements with an infinite number of devices);



- Is used to describe a group of similar hosts service are queuing type G / G / 1 / ∞, which use the service discipline FCFS.

B. *Description node network model*

The network nodes represent the service:

P_p - The time interval between successive P_{xp}-frames, $x_p \in S_P$, $p = \overline{1, L_P}$;

B_b - The time interval between successive B_{xb}-frames, $x_b \in S_B$, $b = \overline{1, L_B}$;

V_v - The time interval between successive V_{xv}-packets of audio accompaniments, $x_v \in S_V$, $v = \overline{1, L_V}$;

D^{Sv} - Delays of sending video server TCP-segments, in which encapsulated RTP-frames with video frames or sound packets;

D_i^{Sw} - Delays of IP-packets switching in the Sw_i-switch; $i = \overline{1, N}$;

D^{Ws} - Delays of IP-packets processing that contain footage taken by RTP in by the Working Station Ws;

D_i^{dw} - Delays in the transmission of Ln_i^G Gigabit Ethernet MAC frames with link-governmental encapsulated IP packets transmitted workstation Ws, $i = \overline{1, N}$;

D_i^{up} - Delays in the transmission of Ln_i^G Gigabit Ethernet MAC frames with link-governmental encapsulated IP packets transmitted to the video server Sw, $i = \overline{1, N}$;

D_i^{pr} - Delays of signal propagation of Ln_i^G Gigabit Ethernet link, $i = \overline{1, N}$;

D^{dw} - Delays of MAC-frames transmission with encapsulated IP packets destined for video server to the workstation in Ln^F the Fast Ethernet link;

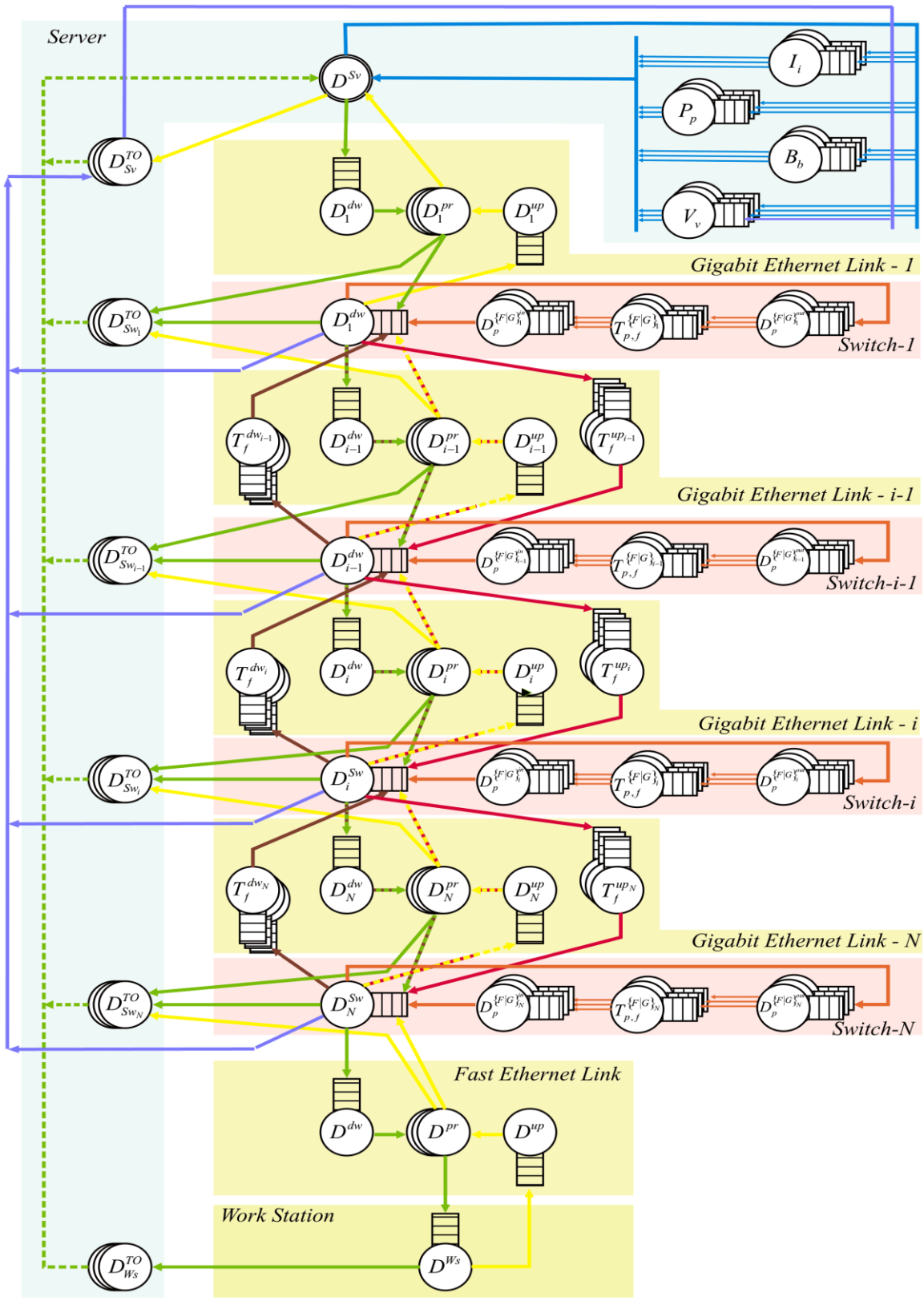


Figure 2. The model structure is the transfer of video traffic over an IP network.

D^{up} - Delays of MAC-frames transmission with encapsulated IP packets transmitted to a video server, Ln^F Fast Ethernet links;

D^{pr} - Delays of signal propagation Ln^F Fast Ethernet links;

D_{Sv}^{TO} - Delays retransmissions of TCP-segments video server due to errors receive IP-packets in a video server Sv;

$D_{Sw_i}^{TO}$ - Delays retransmissions of TCP-segments video server due to errors receive IP-packet switch Sw_i, $i = \overline{1, N}$;

D_{Ws}^{TO} - Delays retransmissions of TCP-segments video server due to errors in receiving IP packets workstation Ws;

$D_p^{F_i^{in}}$ - Delays receiving Sw_i switch MAC-frames received via its input Fast Ethernet port number p. Traffic of these MAC-frames due background traffic; Λ_i^F, $p = \overline{1, N_i^F}$, $i = \overline{1, N}$

$D_p^{G_i^{in}}$ - Delays of MAC-frames receiving Swi switch coming through his incoming Gigabit Ethernet port number p. Traffic of these MAC-frames due background traffic; Λ_i^G, $p = \overline{1, N_i^G}$, $i = \overline{1, N}$

$D_p^{F_i^{out}}$ - Delays of MAC-frame transmission Sw_i Switch through its outbound Fast Ethernet port number p. Traffic of these frames is due to the incoming background traffic; $\Lambda_i^G \;\; u \;\; \Lambda_i^F$, $p = \overline{1, N_i^F}$, $i = \overline{1, N}$

$D_p^{G_i^{out}}$ - Delays of MAC-frame transmission Sw_i Switch through its outgoing Gigabit Ethernet port number p. Traffic of these frames is due to the incoming background traffic; $\Lambda_i^G \;\; u \;\; \Lambda_i^F$, $p = \overline{1, N_i^G}$, $i = \overline{1, N}$

$T_f^{up_i}$ - The time interval between two consecutive data blocks background traffic $\Lambda_{x_f}^{up_i}$, the average size of which is equal to x_f bytes; $x_f \in S_i^{up}$, $f = \overline{1, L_i^{up}}$, $i = \overline{1, N-1}$

$T_f^{dw_i}$ - The time interval between two consecutive data blocks background traffic $\Lambda_{x_f}^{dw_i}$, the average size of which is equal to x_f bytes; $x_f \in S_i^{dw}$, $f = \overline{1, L_i^{dw}}$, $i = \overline{1, N-1}$

$T_f^{G_i}$ - The time interval between two consecutive data blocks background traffic $\Lambda_{x_f}^{G_i}$, the average size of which is equal to x_f bytes; $x_f \in S_i^G$, $f = \overline{1, L_i^G}$, $i = \overline{1, N}$.

$T_f^{F_i}$ - The time interval between two consecutive data blocks background traffic $\Lambda_{x_f}^{F_i}$, whose average size is x_f bytes; $x_f \in S_i^F$, $f = \overline{1, L_i^F}$, $i = \overline{1, N}$.

C. Description of classes network model

Requirement class network model displays a set of similar items of information flows that are transmitted over the telecommunications system being modeled. These elements are, for example, the video frame of the appropriate type and size, data block background traffic, TCP segment, IP packet or MAC frame that encapsulates the appropriate video frame or data block background traffic, a receipt for the corresponding TCP segment. The transfer of items of information-tional flow, according to the protocol used, these elements can change their class.

Classes of requirements are:

I_i^0 - To display the I_{xi} frame in the video stream, the average size of which is equal to x_i bytes, or IP packet (MAC frame) encapsulated in it that I frame, while it was not the case of retransmissions; $x_i \in S_I$, $i = \overline{1, L_I}$

P_p^0 - To display the P_{xp} frame in the video stream, which is equal to the average size of x_p bytes, or IP packet (MAC frame) encapsulated in it that P frame, while it was not the case of retransmissions; $x_p \in S_P$, $p = \overline{1, L_P}$

B_b^0 - To display the B_{xb} frame in the video stream, which is equal to the average size of bytes, or IP packet (MAC frame) encapsulated in it this B frame, while it was not the case of retransmissions; $x_b \in S_B$, $b = \overline{1, L_B}$

I_i^j - To display the IP packet (MAC frame) encapsulated I_{xi} therein-frame, the average size of which is equal x_i bytes, this involved a j its retransmission; $j = \overline{1, J(x_i)}$, $x_i \in S_I$, $i = \overline{1, L_I}$

P_p^j - To display the IP packet (MAC frame) encapsulated P_{xp} therein-frame, the average size of which is equal x_p bytes, this involved a j its retransmission; $j = \overline{1, J(x_p)}$, $x_p \in S_P$, $p = \overline{1, L_P}$

B_b^j - To display the IP packet (MAC frame) encapsulated B_{xb} therein-frame, the average size of which is equal x_b bytes, this involved a j its retransmissions. $j = \overline{1, J(x_b)}$, $x_b \in S_B$, $b = \overline{1, L_B}$

I_i^d - To display the IP packet (MAC frame) encapsulated therein, a I_{x_i} frame, which is equal to the average size of x_i bytes, while the frame itself is brought to the workstation; $x_i \in S_I$, $i = \overline{1, L_I}$

B_b^d - To display the IP packet (MAC frame) encapsulated therein, a B_{x_b} frame, which is equal to the average size of x_b bytes, while the frame itself is brought to the workstation; $x_b \in S_B$, $b = \overline{1, L_B}$

P_p^d - To display the IP packet (MAC frame) encapsulated therein, a P_{x_p} frame, which is equal to the

average size of x_p bytes, while the frame itself is brought to the workstation; $x_p \in S_P,\ p = \overline{1, L_P}$

$\dot{I}_i^j$ - To display the IP packet (MAC frame) encapsulated therein TCP segment, which is delivered to the receipt-I_{x_i} frame, the average size of which is equal x_i bytes, this involved a j retransmissions; $j = \overline{1, J(x_i)},\ x_i \in S_I,\ i = \overline{1, L_I}$

$\dot{B}_b^j$ - To display the IP packet (MAC frame) encapsulated therein TCP segment, which is delivered to the B_{x_b} receipt-frame, the average size of which is equal x_b bytes, this involved a j retransmissions; $j = \overline{1, J(x_b)},\ x_b \in S_B,\ b = \overline{1, L_B}$

$\dot{P}_p^j$ - To display the IP packet (MAC frame) encapsulated therein TCP segment, which is delivered to the P_{x_p} receipt-frame, the average size of which is x_p equal bytes, this involved a j retransmissions; $j = \overline{1, J(x_p)},\ x_p \in S_P,\ p = \overline{1, L_P}$

$\dot{I}_i^d$ - To display the IP packet (MAC frame) encapsulated therein TCP segment, which is delivered on the receipt, but not acknowledged I_{x_i} -frame, the average is equal to the average size of x_i bytes; $x_i \in S_I,\ i = \overline{1, L_I}$

$\dot{B}_b^d$ - To display the IP packet (MAC frame) encapsulated therein TCP segment that is a receipt for delivered but not acknowledged- B_{x_b} frame, the average is equal to the average size of x_b bytes $x_b \in S_B,\ b = \overline{1, L_B}$;

$\dot{P}_p^d$ - To display the IP packet (MAC frame) encapsulated therein TCP segment, which is delivered on the receipt, but not acknowledged- P_{x_p} frame The mean size equal xp bytes $x_p \in S_P,\ p = \overline{1, L_P}$;

$T_{p,f}^{G_i}$ - To display the IP packet (MAC frame) encapsulated therein background data block traffic $\Lambda_{p,x_f}^{G_i}$ coming into the switch Sw_i via its Gigabit Ethernet port number $p, x_f \in S_p^{G_i}, p = \overline{1, N_i^G}, f = \overline{1, L_p^{G_i}}, i = \overline{1, N}$;

$T_{p,f}^{F_i}$ - To display the IP packet (MAC frame) encapsulated therein background data block traffic $\Lambda_{p,x_f}^{F_i}$ coming into the switch Sw_i through its Fast Ethernet port number $p, x_f \in S_p^{F_i}, p = \overline{1, N_i^F}, f = \overline{1, L_p^{F_i}}, i = \overline{1, N}$;

$T_f^{up_i}$ - To display the IP packet (MAC frame) encapsulated therein a block of data background traffic sent from the switch to the switch through the communication Ln_i^G , $x_f \in S^{up_i}, f = \overline{1, L_i^{up}}, i = \overline{2, N}$;

$T_f^{dw_i}$ - To display the IP packet (MAC frame) encapsulated therein data block background traffic sent from the switch to the switch through the communication Ln_i^G , $x_f \in S^{dw_i}, f = \overline{1, L_i^{dw}},\ i = \overline{2, N}$.

Transitions requirements from class I_i^j to class $I_i^{j+1}, j = \overline{0, J(x_i)-1}$, start retransmission model I_{x_i} - frame after receiving an error in the appropriate communication equipment.

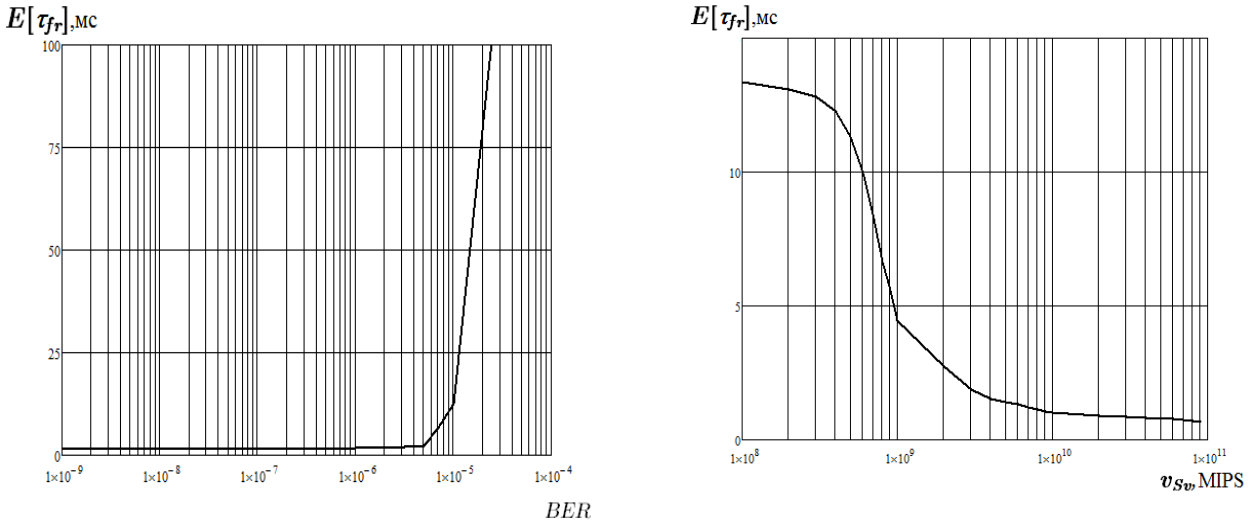
IV. RESULTS OF SIMULATION

For the mathematical modeling of multimedia data transmission via IP-network programming environment Mathcad has been used. Figures 3 and 4 show the results of modeling - quality of service parameters in the transmission of inhomogeneous multimedia traffic in a network of TCP/IP, depending on the characteristics of the communication equipment.

The dependences 2 (a) and 3 (a) show that increasing the probability of bit errors in the channel Gigabit Ethernet to a value of 10-5 of the transmission characteristics (one-way delay and delay variation - jitter) become worse and its quality becomes unacceptable.

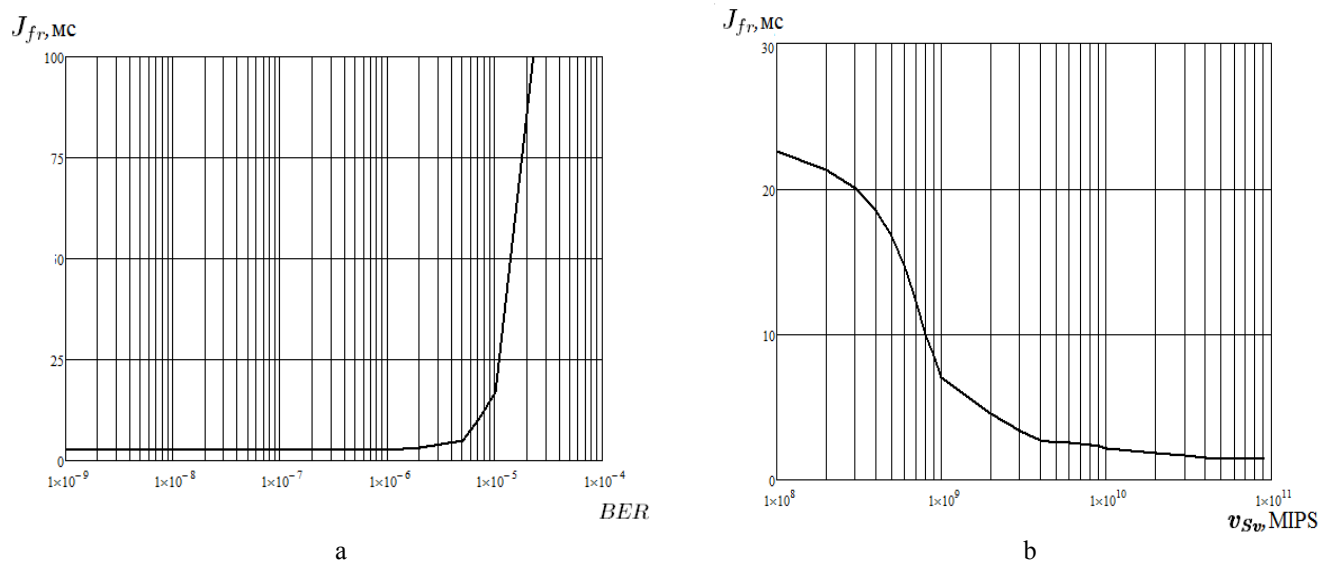
Figures 2 (b) and 3 (b), we can see how are reducing one-way delay and jitter with the increase in server performance. From these graphs it is clear that in order to ensure the required quality of service Video on Demand, you must use a sufficiently powerful computer systems that can quickly handle requests.

Thus, the data can be used to choose the right equipment to provide high-quality transmission video stream from the server to the customers



a b

Figure 3. The dependence of the one-way delay: a) - on the intensity bit error rate; b) - on the performance of the server.



a b

Figure 4. Dependence of jitter: a) - the intensity of the bit error rate, b) - the performance of the server

V. CONCLUSION

In this paper, a method of mathematical modeling of communication networks is proposed in order to estimate the quality of service that is based on the use of inhomogeneous closed queuing networks of large dimensionality. The method is to display different information flows in a simulated telecommunications system by different classes of applications and different delays of data elements relevant information flow queuing. The possible combinations of the delay elements of the information flows are defined applications possible transitions corresponding classes.

Probabilistic-time characteristics of the simulated telecommunications system we calculated according to the expressions for the delays developed in inhomogeneous closed queuing networks [13]. To calculate the parameters of queuing networks the method of analysis medium is applied [13-14].

As an example, a mathematical model was developed to estimate the QoS parameters of transmission inhomogeneous multimedia traffic in IP networks. For this case data one-way delay and jitter were obtained.

The results showed the possibility and high efficiency of applications inhomogeneous closed queuing networks of large dimensionality as a method of mathematical modeling of different communication systems.

REFERENCES

[1] Servetto S.D., Nahrstedt K., Broadcast Quality Video over IP. IEEE Transactions on Multimedia, vol. 3, no.1, March 2001, pp. 162-173.

[2] Hong C., T. Yoneda, K. Tanaka, S. Honda, Y. Matsushita, "Multimedia communication networking architecture model for hign speed network" 19th Conference on Local Computer Networks. Proceedings., pp. 257-265, 1994.

[3] A. Srivastava, A. Kumar, A. Singru and K.A. Kamel, "Object oriented simulation of multimedia on demand services over ATM network" Proceedings of the Fourth International Workshop on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems, MASCOTS'96., pp. 128-132, 1996.

[4] J.-G. Kim, J.P. Hong, B.-J. Min and M.H. Kim, "Multimedia service object modeling" Third International Workshop on Object-Oriented Real-Time Dependable Systems, Proceedings., pp. 347-354, 1997.

[5] X.X. Yao and O.W.W. Yang, "A queueing analysis of slotted ALOHA systems" Canadian Conference on Electrical and Computer Engineering, pp. 1234-1238 vol.2, 1993.

[6] A. Burrell, D Makrakis and P. Papantoni-Kazakos, "Performance analysis of a new implementation efficient capacity allocation protocol for multimedia traffic in ATM networks" Proceedings of the Third IEEE International Symposium on High Performance Distributed Computing, pp. 51-59, 1994.

[7] A. Das, "A model for synchronisation and communication of distributed multimedia data streams" Proceedings of IEEE Singapore International Conference on Networks, pp. 622-626, 1995.

[8] R. Cohen and H. Radha, "Streaming fine-grained scalable video over packet-based networks" Global Telecommunications Conference, GLOBECOM'00., IEEE, vol.1, pp. 288-292, 2000.

[9] W.K. Wong, Y. Qian and V.C.M. Leung, "Scheduling for heterogeneous traffic in next generation wireless network" Global Telecommunications Conference, GLOBECOM'00., IEEE, pp. 283-287, 2000.

[10] R.M. Guirguis and S. Mahmoud, "Transmission of real-time multi-layered MPEG-4 video over ATM/ABR service" 2000 IEEE International Conference on Communications, ICC 2000., vol.1, pp. 259-263, 2000.

[11] T. Irnich and P. Stuckmann, "Analytical performance evaluation of Internet access over GPRS and its comparison with simulation results" The 13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, vol.5, pp. 2127-2131, 2002.

[12] ISO/IEC JTC1/SC29/WG11 №2459 Overview of the MPEG-4 standard. 1998.

[13] A.F. Yaroslavtsev, T.-J. Lee, M.Y. Chung, H. Choo, "Performance Analysis of IP-Based Multimedia Communication Networks to Support Video Traffic Computational Science" ICCS 2004, 4th International Conference, Krakow, Poland, June 6-9, 2004, Proceedings, Part 1, Lecture Notes in Computer Science. – Springer Berlin/Heidelberg, 2004.-Vol.3036. - pp. 573-576.

[14] Chandy K.M., Neuse D., Lineariser: a heuristic algorithm for queueing network models of computing systems // Comm. of the ACM. – 1982. – v.25, No.2.

AvBandTest – a Testing Tool for Implementations of Available Bandwidth Estimation Algorithms

Dmitry Kachan, Eduard Siemens
Anhalt University of Applied Sciences - Department of Electrical, Mechanical and Industrial Engineering
Bernburger Str. 57, 06366 Koethen, Germany
E-mail: d.kachan@emw.hs-anhalt.de, e.siemens@emw.hs-anhalt.de

***Abstract* — This work describes a test tool that allows to make performance tests of different end-to-end available bandwidth estimation algorithms along with their different implementations. The goal of such tests is to find the best-performing algorithm and its implementation and use it in congestion control mechanism for high-performance reliable transport protocols. The main idea of this paper is to describe the options which provide available bandwidth estimation mechanism for high-speed data transport protocols and to develop basic functionality of such test tool with which it will be possible to manage entities of test application on all involved testing hosts, aided by some middleware.**

***Keywords:* High-speed transport; available bandwidth, congestion avoidance, testing of algorihtms.**

I. INTRODUCTION

A transport protocol is a complex system with various number of different logic parts – modules, which together perform transmission of data between involved peers. In [1] we have already shown that even modern commercial transport protocols reveal throughput performance being far from the optimum, and there is a large room for improvements. However, optimization of such a system is a rather complex task, and usually it is reduced to improvement of some particular modules. This paper addresses improvement of congestion avoidance algorithms (congestion control module) and performance of high-speed reliable transport protocols by means of finding a best-suitable available bandwidth estimation mechanism.

II. RELATED WORK

Phenomena like congestion in IP networks occur due to one of the fundamental principle of Internet – best-effort delivery. For reliable transport protocols the only chance to perform well is to use some mechanisms to control instant network utilization and properly react on congestions occurrence. Congestion control mechanisms used in contemporary transport protocols are mostly either window-based or rate-based. Window-based congestion control algorithm are well known and widely used in TCP [2] [3]. Rate-based congestion control has been widely used in ATM systems [4] [3]. However, in [5] Y. Gu et al. use rate based congestion control even in IP networks for a UDP-based transport over high-bandwidth and high-delay links. Their experiments show that the use of rate based congestion control for transport protocols is quite efficient. In [6] L. J. Latecki et al. use slightly modified SLoPS (Self-Loading Periodic Streams) algorithm [7] to develop congestion control for media applications based on estimation of the instantly available bandwidth estimation. A. K. Aggarwal et al. in [8] are using available bandwidth estimation to detect congestions in the data networks. This technique of available bandwidth estimation is also known from TCP [9].

There are many different approaches for measurement of available bandwidth. Most of them are based on PGM (Probe Gab Model) [10] [11] [12] or on PRM (Probe Rate Model) [7]. Within PGM methods, one peer sends a train of packet pairs to a corresponding peer, and based on the dispersion of pairs of packets, receiver peer can make an estimation of the available end-to-end bandwidth. The advantages of these methods are that they are quite fast and generate not much additional traffic in the network. However PGM methods are not able to provide adequate estimation of available bandwidth in presence of cross traffic in the multi-hop path. Considering that Internet has almost always a multi-hop configuration, there are some critical views on the accuracy of results of PGM models for available bandwidth estimation [13].

The methods based on PRM provide more adequate results of estimation. In [14], C. D. Guerrero et al. are comparing common solutions for available bandwidth estimation, and according to this research the minimal value of estimation error has been achieved by *pathload* [7] – a tool based on PRM (SLoPS algorithm). The disadvantages of this method are high estimation time due to multiple iterations of the algorithm, and high load of a link with estimation traffic.

The idea of this work is to use both PGM- and PRM-based algorithms in the ways, where they provide their best advantages: less estimation time or more accurate estimation result to make high speed data transmission more intelligent.

III. AVAILABLE BANDWIDTH ESTIMATION FOR HIGH SPEED TRANSPORT PROTOCOLS

There are two basic ways how to use available bandwidth estimation in transport protocols – initial estimation and estimation during the transmission. First one should be a very

fast method, which, probably, gives not very accurate, but at least approximated values of available bandwidth before the transmission starts. It is necessary to define the initial data rate at the very beginning. For this phase of transmission, an algorithm based on PGM could be used.

Estimation of available bandwidth during transmission is a method, which could take more time, however it will be expected to achieve more precise results. The result of this estimation will be used for soft reaction of transmission on the changes in a network, such as an appearance or disappearance of cross traffic in a path, e.g. by increasing or decreasing of sending rate in order to avoid congestion and for maximal link utilisation. For this kind of estimation PRM-based methods can be used. High speed transport protocols such as UDT [15] or RWTP [16] inject a time stamp into each data packet. It could eliminate the main disadvantage of PRM (high load of a link by probe traffic), since probing traffic could be carried piggy-back in data traffic of the transport protocol and the results could be transmitted by means of ACK messages. It means that no extra traffic will be generated. In that case SLoPS algorithm should be slightly modified to not make active measurements that include sending of probing traffic, but to make a passive, periodically analysing the time stamps in received packets.

To develop and test the modules for available bandwidth estimation it is possible either to implement the respective algorithms within an open source protocol, or to write a light weight application – vehicle, that gives a chance not only to implement algorithms, but also make performance tests to evaluate them. Implementation of such algorithms directly in the source code of a whole protocol stack could unintentionally break the protocol, or modification of one certain software module could negatively affect another module. Besides that testing of such, implemented within the protocol stack, algorithm is difficult because its behaviour would strongly depend also on implementation of the protocol.

IV. Structure and Implementation of AvBandTest

The implementation of the algorithms and accordingly the test tool has been implemented using C++ under Linux operation system. The reason for this it is that the source code of high speed protocols such as UDTv4 [15] or RBUDP [17] are implemented in C/C++. According to it, congestion control mechanism implemented in C++ can easily be linked to these protocol implementations in further tests.

In Figure 1 an interconnection model between two computing nodes is shown. For running tests, two types of traffic are used: first one is probing traffic – that are packets, generated by PGM and PRM algorithms to estimate

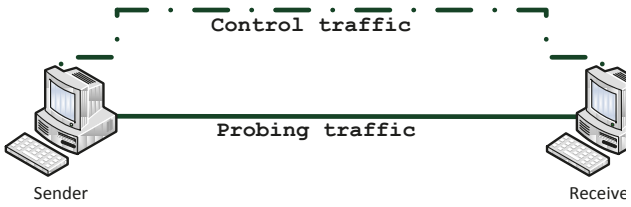


Figure 1. Interconnection scheme

bandwidth; the second one is a control traffic that contains messages which set parameters for available bandwidth estimation, start an estimation session and share result with their peer. Most of the control traffic messages will be used only during experiments to find the best algorithm implementation and the best estimation policy e.g. regarding the amount of probing traffic or the time of estimation, for using it in high speed transport protocols. After implementation of such a mechanism in a real congestion control mechanism, negotiation will be performed by means of service messages of the protocol. It is important that implementation of available bandwidth estimation mechanism will not use additional socket connections.

Since different kinds of impairments take place in telecommunication networks, some packets of both, control and probing traffic, could be received corrupted. The implementation of estimation mechanism should be able to handle corruption of probing traffic. In contrary, integrity of control traffic is out of scope of this test tool and according to that control interconnection will be implemented by means of an existing middleware. Such implementation of interconnection should be robust and allow focusing on the probing traffic only without paying much attention on generating of control information.

The structure of AvBandTest is presented in Figure 2.

A. Middleware

The Object Management Group (OMG) has defined a Common Object Request Broker Architecture (CORBA) [18] decades ago. Different vendors made different implementations of CORBA such as: MICO [19], OmniORB [20] etc. The main problem of CORBA is that the standard was not fully defined, and so different implementations of CORBA are mutually not fully compatible. It made the idea of a really interoperable middleware based on CORBA utopic. Furthermore, the development of many versions of CORBA is already discontinued, e.g. the recent version of MICO has been released September, 14th 2008. Moreover, with definition of a new C++ standard – C++11, applications, that worked under MICO are not compiling anymore. The better

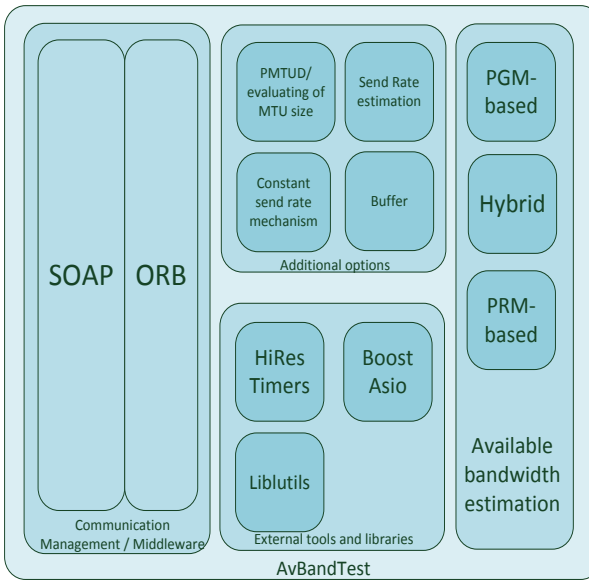


Figure 2. Structure view of AvBand

situation is with OmniORB – latest release was in July of 2011. In contrary to MICO CORBA it allows establishing of connections directly from application, OmniORB uses NameService for communication between hosts. It makes using of such approach not comfortable, because it is needed to starts NameService on each involved host. Moreover CORBA by itself is a sophisticated system with huge amount of its own abstractions, what makes development using this middleware relatively difficult task.

Another widely used protocol specification for remote object invocation is SOAP [21] (Simple Object Access Protocol) which provides a simple and robust technique for message negotiation and data structure exchange. Many different implementations of SOAP use various protocols e.g. HTTP and SMTP for message transmission [22]. There is an open source software development toolkit for C++ – gSOAP, which provides means to automatically generate XML and WSDL code from C++ data and vice versa [23].

On the one hand CORBA has a complete ORB architecture, but difficult enough implementation and not finally defined standard. On the other hand SOAP provides stable communication means, however it has no built-in ORB architecture. The problem of communication management for *AvBandTest* has been solved by development of a simple ORB architecture like one provided by SOAP.

B. External tools and libraries

In C++, all operations with sockets are used from the standard C library. A simple IP communication application, using native C operations, seems quite heavy and hard to read, especially due to the lack of strong type-safety. Moreover, the error handling is also inherited from C, what is pretty inconvenient for C++ programming. Object oriented approach in C++ for handling of IP communications is already implemented e.g. by *Asio* library that is included in a set of libraries called Boost. There is a big community that discusses, makes changes and tests the source code of the Boost library. Moreover, many of Boost libraries in the past have been assigned as C++ standard. The library *Boost.Asio* [24] provides mostly all what is needed to make a comfortable interconnection between sockets including convenient error handling mechanism.

The library *Liblutils* is written by E. Siemens in 2002 and is used within *LTest* [25]. It contains a lot of different function including C++ socket interconnection, but this part is here exchanged to *Boost.Asio* due to the wide supporters community of Boost. *AvBandTest* uses *Liblutils* now only due to a number of convenient strings manipulation functions in C++.

In such tasks as congestion control, time measurement, accurate time fetching and efficient time calculation becomes critical part. In [26] the authors share some novel ideas how to measure the time with the maximal possible resolution on common PC systems by performing of assembly code, wrapped around by C++ interfaces to get an access directly to timer hardware. All these ideas and a number of comfortable functions are implemented in the work of I. Fedotova in the library *HiResTimers* which is also described in [26]. This library is used by *AvBandTest* for time interval measurements.

C. Additional options

1) Buffer

Estimation of available bandwidth will be performed by the transmission of probe traffic on one side and reception of it at another side. For both, PGM and PRM, it is necessary to transmit time stamps in each packet of probe traffic. Beside the time stamp, each sample must carry a sequence number to pervert packets reordering in the network. Both, time stamp and sequence number allocate together not more than 12 bytes of memory: 4 bytes for nanoseconds, 4 bytes for seconds of Unix epoch and 4 bytes for the sequence number. In [11] is shown that the best size for sample packets for PGM should be "not very small and not very large" - between 600 bytes and 1 500 bytes. However it is important to note, that this research has been done with a presumption, that maximal transfer unit (MTU) in internet is 1 500 Bytes. In high speed networks, to achieve maximum capacity, sometimes extended MTU size (Jumbo frames till 9 000 bytes) is used. So it makes sense to check whether these packet sizes are also optimal for extended MTU sizes in the path, or the size of probing packets should be also extended. Nevertheless, there are only 12 bytes of useful data in each sample and to satisfy size conditions of probe packets, the rest space will be allocated with dummy data, filled up randomly.

In a transport protocol the very important part of implementation is the data buffer at the sender and receiver side. The rate of data reception and speed of access to data strongly depends on the buffer implementation. For test purposes *AvBandTool* should also have buffer mechanism that does not slow down the data transmission performance. The buffer has been designed as one separate class, "*ReceivedData*" that contains a vector with the measurement data from received packets. There is no necessity to store whole samples because most of data in the packet do not carry measurement information. For fast storing measured data at the receiver, receiver side will be notified by the sender via the control channel about the amount of expected packets to allocate memory for all expected packets before the reception of IP packets starts. Such operation will be repeated for each measurement's iteration and after calculating of result the buffer will be released. Further improvement of the receiver's buffer could be done by adapting measurements to periodical on-the-fly available bandwidth estimation. This measurement will be continuous with some intermediate results and it will be hard to release the buffer after some certain time period in that case. This problem could be solved by implementation of "*Ring buffer*", which idea is described by E. Siemens et al. in [27].

Beside this vector, the class contains a number of functions to access to particular data components of a sample such as sequence number or time at which the sample has been sent or received. The class has interfaces to calculate and to return such statistical parameters as mean of inter-packet time at the reception or inter-packet time at transmission.

2) Constant send rate mechanism.

Such protocols as RBUDP have an option to send data with some certain data rate that should be chosen once before data transmission starts. However it would be more efficiently to use a mechanism that analyses the actual situation in the

network and notifies the transport engine about changes of available bandwidth. In that case transport will use available resources by maximum.

In *AvBandTest* a rate control mechanism is introduced to emulate data transmission on the data rates that are adopted according the available bandwidth during the data transmission. The main hitch of rate control on the data rates close to 10 Gbit/s is the accurate time measurement. The simple example of transmission of 1 500 bytes packet through 10 Gbit/s network can show it. The mean inter-packet time at the sending side in this case is:

$$t[s] = \frac{S[Byte]}{R\left[\frac{Byte}{s}\right]} = \frac{1\,500}{1\,250 \cdot 10^6} = 1{,}2 \cdot 10^{-6} s, \quad (1)$$

where t – is mean inter-packet time, S – IP packet size and R – used data rate.

Packets smaller than 1 500 has even less inter-packet time, so the measurement tool should be able to measure such short time intervals accurately. In [28] authors show that usual timers in Linux are not stable in choosing of timer source. Furthermore, time request's cost is significantly lower if to request timer system directly, instead of using standard Linux system call "*clock_gettime*". Time measurement by means of direct requests to timers are implemented in [26], where is shown that cost of time requests using this library for different machines were about 1 µs in the worst case and the values of tens of nanoseconds in regular cases. The accuracy of time measurement will always depend on type and hardware realisation of timer. However within *HiResTimer* library it is possible to achieve the most accurate time measurements.

V. First Results

Using Apposite 10G in the 10G-lab of Anhalt University of Applied Sciences (Koethen, Germany), it is possible to emulate IP connections with different bottlenecks on the both sides: sender and receiver. In this way, the accuracy of bandwidth estimation of a prototype version of AvBandTest tool can be tested. The simplified diagram of the testbed is presented in Figure 3. Within presented setup the following bottlenecks were emulated for both sides: sender and receiver: 10 Mbps, 100 Mbps, 1 000 Mbps, 10 000 Mbps. At the time of writing of this paper, the only easiest PGM-based algorithm has been implemented. It works as follow: sender sends to the receiver a predefined amount of packets, which can be defined at the time of application start. These packets have been sent back-to-back – with minimal possible delay between them. Each packet contains: sequence number of a packet, time stamp, as described in section IV. Sender evaluates sending data rate, while it sends data, as sum of bytes of first thousand packets, or even less, if the whole transmission contains less than thousand packets, and divides it by time, that sender spent on sending of this amount of data. After transmission the

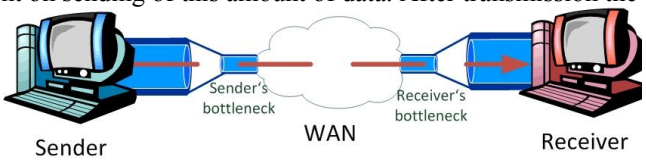


Figure 3 Simplified chart of testbed topology

receiver calculates the differences between timestamps that were included in each two consecutive data packets and calculates the mean value of these differences. For the same pairs of packets receiver calculates the mean value of time differences between the moments of receiving of each packet. The available bandwidth is evaluated as shown in (2)

$$AvBand[bit/s] = \frac{R_s[bit/s]}{\frac{1}{n}\sum_{i=1}^{n}\frac{\Delta t_i^R[s]}{\Delta t_i[s]}}, \quad (2)$$

where $AvBand$ is available bandwidth; R_s is sending rate; n – is a whole number of received pairs; Δt_i^R – is an inter-packet time at reception for i pair of packets; Δt_i – is an inter-packet time at transmission for i pair of packets.

Practically this method means that available bandwidth is back-proportional to a relation of sending inter-packet time interval and reception inter-packet time interval.

The plot in Figure 4 shows available bandwidth estimation error for each emulated bottleneck. The error here is a difference between emulated bottleneck and result of estimation in percent from bottleneck. Figure 5 shows the results of evaluation on a link with 50 ms of RTT, uniformly distributed in the forward and backward directions, and 0.5% of packet losses in the network in between the sender and receiver. "s" and "r" on the plots correspond to the positioning of the bottleneck on sender side and on the receiver side accordingly. In the case of 10 000 Mbps there are no bottlenecks on the sender and receiver sides. The evaluation has been performed 5 times for each combination of

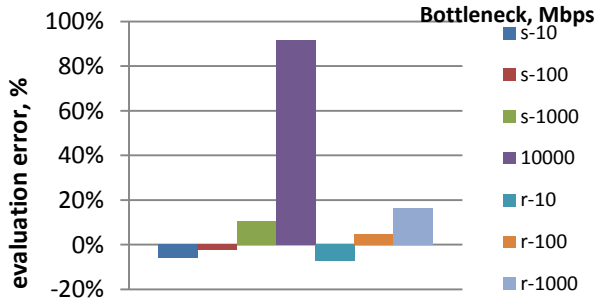


Figure 4. Error of available bandwidth evaluation in the network without impairments

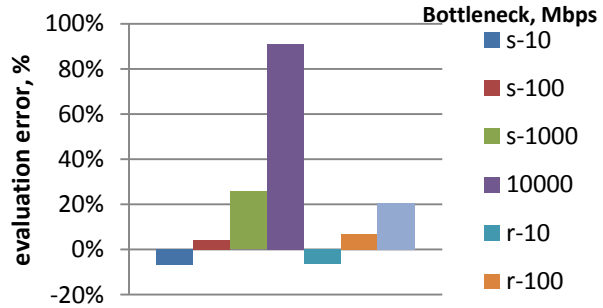


Figure 5. Error of available bandwidth evaluation in presence of 50 ms. of RTT and 0,5% of packet loss in the network

bottlenecks and the average values of mistakes are shown. It is worth noting that within bottleneck on sender side the result of evaluation is better in the network without impairments; however, in presence of them this behavior is not saved. We assume that for initial available bandwidth measurement the error of about 15% is acceptable. The obtained results showed that algorithm needs improvement. The current prototype implementation of AvBandTest is not able to measure available bandwidth up to 10 000 Mbps: as the plots show, in both cases evaluation of 10 000 was completely wrong. Improvement of this point is a significant task for further work.

VI. Conclusion

This work describes problems of available bandwidth estimation tests, which are used for development of fast data transport protocols. There are two significant parameters which can be tested with such techniques: estimation of the very beginning sending rate and estimation of the target data rate of rate based congestion control. Discussion about basic functionality and brief overview of components are presented in this work. A prototype of the tool that allows evaluation of different approaches for available bandwidth measurements has been implemented and tested as a result of this work.

VII. References

[1] D. Kachan, E. Siemens, V. Shuvalov. Comparison of contemporary solutions for high speed data transport on WAN connections. Accepted to: ICNS, Lisbon, Portugal. 2013.

[2] M. Allman, V. Paxson, W. Stevens. M. Allman, V. Paxson, and W. Stevens. IETF RFC 2581. 1999.

[3] A. S. Tanenbaum. Computer Networks. Third edition. New Jersey : Prenice Hall PRT, 1996. p. 813. ISBN 0-13-349945-6.

[4] H. Ohsaki et al. Rate-based congestion control for ATM networks. ACM SIGCOMM Computer Communication Review. 1995, Vol. 25, pp. 60-72.

[5] Y. Gu, X. Hong, M. Muzzucco, R. Grossman. Rate Based Congestion Control over High Bandwidth/Delay Links. IEEE/ACM Transaction on Networking. 2003, Vol. 11.

[6] L. J. Latecki, T Jin, J. Mulik. A Two-stream Approach for Adaptive Rate Control in Multimedia Applications. IEEE Int. Conf. on Multimedia & Expo. 2004.

[7] M. Jain, C. Dovrolis. End-to-end available bandwidth: Measurement methodology, dynamics, and relation with TCP throughput. IEEE/ACM Transactions on Networking. 2003, Vol. 11 (4), pp. 537-549.

[8] A. K. Aggarwal, A. N. Bharadwaj, R.D. Kent. Active congestion control using available bandwidth-based congestion detection. Proc. ICAI'05/MCBC'05/AMTA'05/MCBE'05. 2005, pp. 390-397.

[9] R. Wang et al. Efficiency/Friendliness Tradeoffs in TCP Westwood. Proc. Computers and Communications, ISCC. 2002, pp. 304-311.

[10] C. Dovrolis, P. Ramanathan, D. Moor. What do packet dispersion techniques measure? proceedings of INFOCOM 2001. Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies. 2001, Vol. 2, pp. 905-914.

[11] M. Jain C. Dovrolis. Packet-Dispersion Techniques and a Capacity-Estimation Methodology. IEEE/ACM Transaction on Networking. 2004, Vol. 12, pp. 963-977.

[12] V. J. Ribeiro et al. pathChirp: Efficient Available Bandwidth Estimation for Network Paths. Passive and Active Measurement Workshop. 2003.

[13] L. Lao, C. Dovrolis, M. Y. Sanadidi. The probe gap model can underestimate the available bandwidth of multihop paths. ACM SIGCOMM Computer Communication Review. 2006, Vol. 36 issue5.

[14] C. D. Guerrero, M. A. Labrador. On the applicability of available bandwidth estimation techniques and tools. Computer Communications. Vol. 33 Issue 1, 2010, pp. 11-22.

[15] Y. Gu, R. L. Grossman. UDT: UDP-based Data Transfer for High-Speed Wide Area Networks. Computer Networks (Elsevier). May 2007.

[16] E. Siemens, R. Einhorn, A. Aust. Multi-Gigabit Challenges: Similarities between Scientific Environments and Media Production. ACIT - Information and Communication Technology : s.n., 2010.

[17] E. He, J. Leigh, O. Yu, T. DeFanti. Reliable Blast UDP: Predictable High Performance Bulk Data Transfer. In Proceeding of 5th Int. Conf. on Cluster Computing. 2002.

[18] S. Vinoski. CORBA: integrating diverse applications within distributed heterogeneous environments. Communications Magazine, IEEE. 1997, Vol. 35, pp. 46-55.

[19] A. Puder, K. Römer. MICO: An Open Source CORBA Implementation. - 3rd ed. s.l. : Morgan Kaufmann Publishers, 2000.

[20] S.-I. Lo, S. Pope. The Implementation of a High Performance ORB over Multiple Network Transports. In Middleware 98: IFIP International Conference on Distributed Systems Platforms and Open Distributed Processing. 1998.

[21] M. Gudgin et al. W3C Recomendations. SOAP Version 1.2 Part 1: Messaging Framework (Second Edition). [Online] 04 27, 2007. [Cited: 01 2013, 14.] http://www.w3.org/TR/soap12-part1/.

[22] F. Curbera. Unraveling the Web services web: an introduction to SOAP, WSDL, and UDDI . Internet Computing, IEEE. 2002, Vols. 6 , Issue: 2 , pp. 86-93.

[23] R. A. van Engelen. gSOAP Toolkit. The gSOAP Toolkit for SOAP Web Services and XML-Based Applications . [Online] [Cited: 01 15, 2013.] http://www.cs.fsu.edu/~engelen/soap.html.

[24] C. Kohlhoff . Boost C++ Libraries. Boost.Asio. [Online] [Cited: 01 15, 2013.] http://www.boost.org/doc/libs/1_52_0/doc/html/boost_asio.html.

[25] E. Siemens, S. Piger, C. Grimm, M. Fromme. LTest – A Tool for Distributed Network Performance Measurement. Proc. Consumer Communications and Networking Conference, 2004. First IEEE. 2004, pp. 239-244.

[26] I. Fedotova, E. Siemens, H. Hu. A High-precision Time Handling Library for Tracking Internet Packet Dynamics. Accepted to: ICNS, Lisbon, Portugal. 2013.

[27] E. Siemens, Xiaopeng Qiu. Datentransport in multimedialen Systemen: Effiziente Pufferspeicher für schnellen Datentransport. 2009. 978-3836479837.

[28] D. Kachan, E. Siemens, H. Hu. Tools for the high-accuracy time measurement in computer systems. T-Comm – Telecommunication and Transport. 2012, 8, pp 23-27.